Table of Values

Value and Units	*Item*		*Symbol or Abbreviation*	*Derivation of Value*
• *General*				
57.3 degrees (57°20′)	$\cong$ 1 radian			$180°/\pi$
3.44×10^{3} minutes (of arc)	$\cong$ 1 radian			
2.06×10^{5} seconds (of arc)	$\cong$ 1 radian			
0.0174 radian	$\cong$ 1 degree		$^{\circ}$	$\pi/180°$
2.91×10^{-4} radian	$\cong$ 1 minute (of arc)		$'$	
4.85×10^{-6} radian	$\cong$ 1 second (of arc)		$''$	
1.61×10^{5} cm	$\cong$ 1 statute mile			
10^{-8} cm	$\equiv$ 1 angstrom		Å	
10^{-4} cm	$\equiv$ 1 micron		μ	
2.998×10^{2} volts	$\cong$ 1 statvolt			$10^{-8} c$
2.998×10^{10} cm/sec	Speed of light in vacuum		c	
≈ 980 cm/sec^2	Acceleration of gravity at earth's surface		g	GM_E/R_E^2
6.670×10^{-8} dyne-cm^2/gm^2	Gravitational constant		G	
1 dyne	$\equiv$ 1 gm cm/sec^2			
• *Astronomical*				
3.084×10^{18} cm	$\cong$ 1 parsec			
9.46×10^{17} cm	$\cong$ 1 light-year		*ly*	$c \times$ sec/yr
1.49×10^{13} cm	$\cong$ 1 astronomical unit ($\equiv$ radius of earth's orbit)		A.U.	
$\sim 10^{80}$	Number of nucleons	Known universe		
$\approx 10^{28}$ cm	Radius			
$\sim 10^{11}$	Number of galaxies			
$\approx 3 \times 10^{-18}$ (cm/sec)/cm	Speed of recession of nebulae			
$\approx 1.6 \times 10^{11}$	Number of stars	Galaxy		
$\sim 10^{23}$ cm	Diameter			
$\sim 8 \times 10^{44}$ gm	Mass			
6.96×10^{10} cm	Radius	Sun		
2.14×10^{6} sec	Period of rotation			
1.99×10^{33} gm	Mass			
1.49×10^{13} cm	Radius of orbit	Earth		
6.37×10^{8} cm	Mean radius			
5.98×10^{27} gm	Mass			
5.52 gm/cm^3	Mean density			
3.16×10^{7} sec	$\cong$ 1 year (period of revolution)			
8.64×10^{4} sec	24 hours (period of rotation)			
3.84×10^{10} cm	Radius of orbit	Moon		
1.74×10^{8} cm	Radius			
7.34×10^{25} gm	Mass			
2.36×10^{6} sec	Period of revolution			

TATA McGRAW-HILL PUBLISHING COMPANY LTD.

BOMBAY — NEW DELHI

mechanics

berkeley physics course — volume 1

The preparation of this course was supported by a grant from the National Science Foundation to Educational Services Incorporated.

8753
West Bengal
STATE INSTITUTE OF EDUCATION

Charles Kittel

Professor of Physics
University of California
Berkeley

Walter D. Knight

Professor of Physics
University of California
Berkeley

Malvin A. Ruderman

Professor of Physics
New York University

Library
Calcutta
23/3, B. C. Road

MECHANICS

Copyright © 1962, 1963, 1964, 1965 by Education Development
Center, Inc. (successor by merger to Educational Services
Incorporated).

All Rights Reserved.
This book, or parts thereof, may not be reproduced in any form
without the written permission of Education Development
Center Inc., Newton, Massachusetts.

The copyright owner will give permission for use of the
original work in the English language after January 1, 1975.
For conditions of use, permission to use, and for other
permissions, apply to the copyright owner.

[illegible]
Date. 17.2.2005
Accn. No. 10978

T M H Edition

*Reprinted in India by arrangement with McGraw-Hill Book Company, Inc.
New York.*

This edition can be exported from India only by the Publishers,
Tata McGraw-Hill Publishing Company Ltd.

Sales Territory : India, Burma and Ceylon.

This book has been published with the assistance of the
Joint Indian-American Textbook Programme

Price Rs. 10.50

Published by Tata McGraw-Hill Publishing Company Limited and
printed by Mohan Makhijani at Rekha Printers, New Delhi-55

Foreword

One of the urgent problems confronting universities today is that of undergraduate teaching. As research has become more and more absorbing to the faculty, a "subtle discounting of the teaching process" (to quote philosopher Sidney Hook) has too often come into operation. Additionally, in many fields the changing content and structure of knowledge growing out of research have created great need for curriculum revision. This is particularly true, of course, in the physical sciences.

It is a pleasure, therefore, to contribute a foreword to the Berkeley Physics Course and Laboratory, which is a major curriculum improvement program at the undergraduate level designed to reflect the tremendous revolutions in physics of the last hundred years. The course has enlisted the efforts of many physicists working in forefront areas of research, and has been fortunate to have the support of the National Science Foundation, through a grant to Educational Services Incorporated. It has been tested successfully in lower division physics classes at the University of California, Berkeley, over a period of several semesters. The course represents a marked educational advance, and I hope it will be very widely used.

The University of California is happy to act as host to the interuniversity group responsible for developing this new course and laboratory, and pleased that a number of Berkeley students volunteered to help in testing the course. The financial support of the National Science Foundation and the cooperation of Educational Services Incorporated are much appreciated. Most gratifying of all, perhaps, is the lively interest in undergraduate teaching evinced by the substantial number of University of California faculty members participating in the curriculum improvement program. The scholar-teacher tradition is an old and honorable one; the work devoted to this new physics course and laboratory shows that the tradition is still honored at the University of California.

Clark Kerr

8733

Preface to the Berkeley Physics Course

This is a two-year elementary college physics course for students majoring in science and engineering. The intention of the writers has been to present elementary physics as far as possible in the way in which it is used by physicists working on the forefront of their field. We have sought to make a course which would vigorously emphasize the foundations of physics. Our specific objectives were to introduce coherently into an elementary curriculum the ideas of special relativity, of quantum physics, and of statistical physics.

This course is intended for any student who has had a physics course in high school. A mathematics course including the calculus should be taken at the same time as this course.

There are several new college physics courses under development in the United States at this time. The idea of making a new course has come to many physicists, affected by the needs both of the advancement of science and engineering and of the increasing emphasis on science in elementary schools and in high schools. Our own course was conceived in a conversation between Philip Morrison of Cornell University and C. Kittel late in 1961. We were encouraged by John Mays and his colleagues of the National Science Foundation and by Walter C. Michels, then the Chairman of the Commission on College Physics. An informal committee was formed to guide the course through the initial stages. The committee consisted originally of Luis Alvarez, William B. Fretter, Charles Kittel, Walter D. Knight, Philip Morrison, Edward M. Purcell, Malvin A. Ruderman, and Jerrold R. Zacharias. The committee met first in May 1962, in Berkeley; at that time it drew up a provisional outline of an entirely new physics course. Because of heavy obligations of several of the original members, the committee was partially reconstituted in January 1964, and now consists of the undersigned. Contributions of others are acknowledged in the prefaces to the individual volumes.

The provisional outline and its associated spirit were a powerful influence on the course material finally produced. The outline covered in detail the topics and attitudes which we believed should and could be taught to beginning college students of science and engineering. It was never our intention to develop a course limited to honors students or to students with advanced standing. We have sought to present the principles of physics from fresh and unified viewpoints, and parts of the course may therefore seem almost as new to the instructor as to the students.

The five volumes of the course as planned will include:

I. Mechanics (Kittel, Knight, Ruderman)
II. Electricity and Magnetism (Purcell)
III. Waves and Oscillations (Crawford)
IV. Quantum Physics (Wichmann)
V. Statistical Physics (Reif)

The authors of each volume have been free to choose that style and method of presentation which seemed to them appropriate to their subject.

The initial course activity led Alan M. Portis to devise a new elementary physics laboratory, now known as the Berkeley Physics Laboratory. Because the course emphasizes the principles of physics, some teachers may feel that it does not deal sufficiently with experimental physics. The laboratory is rich in important experiments, and is designed to balance the course.

The financial support of the course development was provided by the National Science Foundation, with considerable indirect support by the University of California. The funds were administered by Educational Services Incorporated, a nonprofit organization established to administer curriculum improvement programs. We are particularly indebted to Gilbert Oakley, James Aldrich, and William Jones, all of ESI, for their sympathetic and vigorous support. ESI established in Berkeley an office under the very competent direction of Mrs. Minty R. Maloney to assist the development of the course and the laboratory. The University of California has no official connection with our program, but it has aided us in important ways. For this help we thank in particular two successive Chairmen of the Department of Physics, August C. Helmholz and Burton J. Moyer; the faculty and nonacademic staff of the Department; Donald Coney, and many others in the University. Abraham Olshen gave much help with the early organizational problems.

Your corrections and suggestions will always be welcome.

Eugene D. Commins
Frank S. Crawford, Jr.
Walter D. Knight
Philip Morrison
Alan M. Portis
Edward M. Purcell
Frederick Reif
Malvin A. Ruderman
Eyvind H. Wichmann
Charles Kittel, *Chairman*

Berkeley, California

Preface to Volume I

The subject of this volume of the Berkeley Physics Course is elementary mechanics. Our approach is not radical, but differs perhaps from many textbooks in several aspects:

1. The consequences of the special theory of relativity are developed in detail. The central results here are essential to the development of electricity and magnetism in Vol. II.

2. We have emphasized the motion of charged particles in electric and magnetic fields. This area is rich in simple and important applications, and it relates immediately to the early experiments in the Berkeley Physics Laboratory.

3. We have tried to present elementary mechanics so that paths are seen leading to other parts of physics, to astronomy, to geophysics, and (as far as we could) to chemistry and biophysics. Many astronomical problems and examples are included.

4. We have tried to approach problems as most physicists would approach them, thereby hoping to develop early in the student a facility in some of the unwritten methods of scientific research and reasoning. We have emphasized order-of-magnitude estimates and dimensional analysis.

5. We have provided (in the form of advanced topics at the ends of the chapters) important relevant material for the superior student.

The first version of this volume, written by M. A. Ruderman, was used by an experimental class at Berkeley in the spring of 1963. It was then revised by C. Kittel with the assistance of W. D. Knight. We benefited from criticism by Philip Morrison, Edward M. Purcell, A. C. Helmholz, Allan M. Portis, Eyvind H. Wichmann, David Korff, Bernard Friedman, Allan Kaufman, W. A. Nierenberg, and others. The many drawings which are such a vital part of the volume were created by Eugene D. Commins and were drawn in final form by Felix Cooper. The second version was used at Berkeley and at Maryland by experimental classes in the fall of 1963, and a revision of the second version was used by regular classes of 230 students at Berkeley and 45 students at the University of Texas in the spring of 1964. The enthusiastic student response was enormously encouraging to the exhausted authors. The revised second version was extensively revised again in the summer of 1964.

The entire volume was reviewed by Simon Pasternack; he caused extensive improvements to be made. R. McPherron, H. Ohanian, A. Felzer, R. Kirschman, and others helped with the problems, and Michael Rossman helped write and collect the ancillary material. We have benefited from literary criticism by Thomas Parkinson. For advice on geophysical problems we are indebted to John Verhoogen, Bruce Bolt, and J. H. Reynolds; on astronomical problems to Paul Hodge; and on biophysics to R. C. Williams, G. Stent, W. D. Phillips, and H. K. Schachman. For expert advice on relevant films we wish to thank Robert Hulsizer. Robert R. Davis and others of the staff of *Physics Today* kindly helped locate many photographs. Mrs. G. Titus and Mrs. Kimio Hom have helpfully located many obscure references for us. Extensive and valuable feedback from students and teaching assistants was organized by Charles LeVine during the third trial of the preliminary material. He also reviewed the final text and is responsible for the answers that accompany the problems. J. Ryus did much valuable proofreading. Mrs. Madeline Moore helped organize the manuscript. Many other persons gave occasional assistance.

C. Kittel
W. D. Knight
M. A. Ruderman

Teaching Notes

We have deliberately included more material in Vol. I than is likely to be covered in formal lectures to first-year undergraduates. At Berkeley in 1963 and 1964 we taught the most essential material in this volume in one semester of 15 weeks, with three lectures and one discussion section a week, each of 50 minutes. The students were second-semester freshmen who had had a semester of calculus, not specifically oriented to the needs of students of physics. The following suggestions are derived from our early experience and from extensive student feedback under these conditions.

The terminal boundary condition on the mechanics part of our course is that *the Lorentz transformation of space and time (Chap. 11) and of momentum and energy (Chap. 12) must be treated in full detail as a necessary prerequisite for the development of electricity and magnetism in Vol. II.* In our opinion, *instruction must start on Chap. 10 before two-thirds of the total time has elapsed, no matter what earlier material is omitted to do this.*

Some topics (such as waves and heat) usually treated in the mechanics part of introductory physics have been deferred to Vols. III and V. The advanced topics and mathematical notes are optional, except where otherwise stated. Most students will read the historical notes without prompting.

From our surveys of student response the discussion sections appear to be an unusually important element in the successful instruction of this course. It is possible that two lectures and two recitation sections a week would be more effective than the "three and one" we have used. The lectures should include demonstrations. The students suggest that the lectures should paraphrase, illuminate, and summarize the text rather than repeat derivations word for word.

Examples, Problems, and Advanced Topics

Many *examples* are worked out in the text; these for the most part are essential to the development and are to be read as part of the text. At the end of each chapter are placed the *problems,* which are intended to give the student practice in the application of the principles discussed in the chapter. The easier problems tend to occur at the beginning of each group. Answers are included with a substantial proportion of the problems. The *advanced topics* are arranged to lead the superior student into new areas of physics or otherwise to exploit the methods taught in the chapter.

Mathematical Notes

Several chapters are accompanied by supplementary mathematical material intended to help the student anticipate topics that may arise in his physics course before they are treated in his mathematics courses. We have not hesitated to refer students to integral tables, in an effort to reduce the quantity of mathematics treated in the text.

Historical Notes

Fragmentary historical notes are appended to many of the chapters in Vol. I. Some of the notes are comprised of excerpts from one or two of the great and pioneer original research papers in a particular field of physics. In presenting these excerpts our primary intent is to show the clarity, energy, and courage characteristic of great new contributions and discoveries. The clearest paper in a field is usually the first paper, and it may be the *only* paper to make plain the motivation for trying a new approach. The element of courage, which enters in not being afraid to do something new, plays an exceedingly important part in research. Our secondary intent is to encourage the student to become familiar with the original literature of physics.

Chapter 1. (Introduction) This is fairly light reading. The *advanced topic* on simple astronomy within the solar system has a strong appeal to students; it might be assigned in place of any problems. The lecturer may usefully discuss order-of-magnitude problems ranging from the atom to the universe.

Chapter 2. (Vectors) This chapter is easy to teach. The derivatives of sine and cosine are introduced for use in Chap. 3. Arrange with your mathematics department for these to be taught in time, or do it yourself. Lecture demonstrations are useful, to exhibit the vector addition of velocities and of forces, and to show what torque is.

Chapter 3. (Galilean invariance) In a minimum program the discussion of the Coriolis acceleration as an advanced topic is optional. By treating the special case of a particle at rest in the rotating reference frame one finds the centripetal acceleration, which is utilized later in many places in the volume. The film *Frames of reference* by Hume and Ivey should definitely be shown. A good experiment is to dip a ball bearing in paint and then project it across a rotating platform.

Chapter 4. (Particle dynamics) Although the problems are elementary, this chapter cannot be rushed. It causes some students considerable temporary difficulty, because they are now applying mathematics to physics almost for the first time. A strong argument for teaching this material early is its relevance to the Berkeley Physics Laboratory. *Analogies with motion in a gravitational field should be emphasized,* because these are familiar from high school. The term "field" tends to panic students, just as "prose" does in English classes. In a minimum program in mechanics it is not necessary to deal with complex numbers. Lecture demonstrations may include linear accelerated motion; uniform circular motion; waveforms (using a projection oscilloscope); and the rolling wheel.

Chapter 5. (Energy) The material of this chapter is fairly standard. The distinction between applied and interaction forces may cause some initial confusion which may be resolved by working a number of problems at the blackboard. Demonstrations include springs and pendulums. The concept of line integral may be omitted.

Chapter 6. (Momentum) The collision and satellite problems deserve full discussion. The equations could be set up for Rutherford scattering (solution given in Chap. 15). The astronomical examples will interest the livelier students, but may be omitted from a minimum program. Demonstrations include jet-propelled toy rockets; the ballistic pendulum; the rotating chair.

Chapter 7. (Harmonic oscillator) The linear problems, and particularly the driven harmonic oscillator, are very important. Even in a minimum program the first of the three nonlinear examples should be discussed, because it gives the students confidence to know that they can estimate the errors in the linearized pendulum problem. The idea of phase angle in the response of a driven harmonic oscillator does not come naturally to most students. A good lecture demonstration helps here. The electrical analogs cause confusion at this stage and might perhaps be left to the laboratory. Demonstrations include harmonics in tuning forks (amplify and listen to sound and exhibit waveform on scope); driven spring-mass system; response of driven *LRC* circuit with signal generator; Pringsheim's apparatus; coupled oscillators.

Chapter 8. (Rigid-body dynamics) This chapter should be omitted from a minimum program. Demonstrations include gyros; NMR or ESR; and wheel-and-axle experiments.

Chapter 9. (Central forces) The planetary orbit problem is easy to teach in the form in which it is presented. It is the classical problem of classical mechanics. It should be omitted if two-thirds of the total time has now elapsed. Demonstrations include the Cavendish experiment; mass on string in horizontal orbit of contracting radius; and rotation of a dumbbell model of the earth-moon system to show a dynamic method for locating the center of mass.

Chapter 10. (Speed of light) A chapter largely for private reading. The discussion of the Doppler effect cannot be overemphasized. The problems emphasize the astronomical scene. Demonstrate a velocity of light experiment; if possible, Doppler effect à la Mössbauer. Exhibit an interferometer.

Chapter 11. (Lorentz transformation of length and time) This chapter is of central importance. It is surprisingly easy to teach. Many problems should be assigned. Recapitulate the discussions on coordinate transformations in Chap. 3 (and Chap. 4). Emphasize always the invariance of *c*. Discuss the Ives and Stillwell experiment.

Chapter 12. (Relativistic momentum and energy) The golden harvest of the special theory of relativity is given in Chaps. 12 and 13. The historical note on the mass-energy relation is independent of the chapter and is easily understood. Discuss in class the design of beam deflectors and experimental details of the Bucherer transverse-momentum experiments; show slides of bubble-chamber pictures.

Chapter 13. (Relativistic dynamics) The problems discussed are not difficult, and they help make the student feel at home in the world of special relativity.

Chapter 14. (Equivalence) No particular difficulties. Discuss in class the details of the Pound-Rebka experiment.

Chapter 15. (Particles) A short discussion for reference of the elementary facts about the more important particles.

On examinations: The students will need explicit direction on the topics on which they are likely to be examined; otherwise they may be disconcerted by the quantity of ancillary material presented in the text. Some of the material which should be covered is less familiar and may seem more difficult than that covered in many elementary texts. The examinations can take account of this by emphasizing simple, rather than intricate or tricky, situations, because the object even of an examination is to teach physics and not to be an intelligence test.

Note to the Student

The beginning year of college physics is by far the most difficult. In the first year many more new ideas, concepts, and methods are developed than in advanced undergraduate or graduate courses. A student who understands clearly the basic physics developed in this first volume, even if he may not yet be able to apply it easily to complex situations, has put behind him most of the real difficulties in learning physics.

What should a student do who has difficulty in working the problems and in understanding parts of this course, even after reading the text twice? First he should go back and reread the relevant parts of his high school physics book and he should read deeply in Physics, *the PSSC text. He can usefully consult one of the many books at the beginning college level which are simpler and more elementary than our own; he should look particularly at the exercises which are worked out in other texts. An excellent review of the elements of calculus is available as a short manual of self-instruction:* Quick calculus, *by Daniel Kleppner and Norman Ramsey (John Wiley and Sons, New York, 1965). In a very short time this manual will bring your calculus up from zero to the level needed for our course.*

Notation

Units

Every mature field of science and engineering has its own special units for quantities which occur frequently. The *acre-foot* is a natural unit of volume to an irrigation engineer, a rancher, or an attorney in the western United States. The *Mev* or *million electron volts* is a natural unit of energy to a nuclear physicist; the *kilocalorie* is the chemist's unit of energy, and the *kilowatt-hour* is the power engineer's unit of energy. The theoretical physicist will often simply say: Choose units such that the speed of light is equal to unity. A working scientist does not spend much of his time converting from one system of units to another; he spends much more time in keeping track of factors of two and of plus or minus signs in his calculations Nor will he spend much time in arguing about units, because no good science has ever come out of such an argument.

Physics is carried out and published chiefly in three systems of units, the Gaussian CGS, the MKS, and the practical. Every scientist and engineer who wishes to have easy access to the literature of physics will need to be familiar with all three systems.

The text is written in the Gaussian CGS system. The laboratory makes some use of all three systems. This policy was decided unanimously by our original committee and is in conformity with our object to present physics as most physicists use it. It is clear to anyone who leafs through current physics journals that more physics papers employ the Gaussian CGS system than any other system. In a physics course we want to make it as easy as possible for both scientists and engineers to read physics journals.

Physical Constants

Approximate values of physical constants and useful numerical quantities are printed inside the front and back covers of this volume. More precise values of physical constants are tabulated in *Physics Today*, pp. 48–49, (Feb. 1964).

Signs and Symbols

In general we have tried to adhere to the symbols and unit abbreviations which are used in the physics literature—which are, for the most parr, agreed upon by international convention. In a few cases pedagogical considerations have made it advisable to deviate from these conventions. For example, we use the abbreviations gm and ev respectively, for grams and electron volts whereas the conventional abbreviations are g and eV.

We summarize here several signs which are used freely throughout the course.

$=$ is equal to
$\cong$ is nearly equal to
$\equiv$ is identical with
$\propto$ is proportional to
$\approx$ is approximately equal to; is roughly equal to
$\sim$ is of the order of magnitude of

Usage of the signs $\approx$, $\cong$, and $\sim$ is not standardized, but the definitions we have given are employed fairly widely by physicists. The American Institute of Physics encourages use of the sign $\approx$ where others might write either $\approx$ or $\cong$. (*Style Manual*, American Institute of Physics, rev. ed., March 1963.)

The sign $\sum_{j=1}^{N}$ or $\sum_{j}$ denotes summation over what stands to the right of Σ over all entries between $j = 1$ and $j = N$. The notation $\sum_{i,j}$ denotes double summation over the two indices i and j. The notation $\sum_{i,j}{}'$ or $\sum_{\substack{i,j\\ i\neq j}}$ denotes summation over all values of i and j except $i = j$.

Order of Magnitude

By this phrase we usually mean "within a factor of 10 or so." Free and bold estimation of the order of magnitude of a quantity characterizes the physicist's work and his mode of speech. It is an exceptionally valuable professional habit, although it often troubles beginning students enormously. We say, for example, that 10^4 is the order of magnitude of the numbers 5500 and 25,000. In CGS units the order of magnitude of the mass of the electron is 10^{-27} gm; the accurate value is $(0.91072 \pm 0.00002) \times 10^{-27}$ gm.

We say sometimes that a solution includes (is accurate to) terms of order x^2 or E, whatever the quantity may be. This is also written as $O(x^2)$ or $O(E)$. The language implies that terms in the exact solution which involve higher powers (such as x^3 or E^2) of the quantity may be neglected for certain purposes in comparison with the terms retained in the approximate solution.

Prefixes

The following tabulation shows the abbreviation and numerical significance of some frequently used prefixes:

10^{12}	T	tera-	10^{-3}	m	milli-
10^{9}	G	giga-	10^{-6}	μ	micro-
10^{6}	M	mega-	10^{-9}	n	nano-
10^{3}	k	kilo-	10^{-12}	p	pico-

Greek Alphabet

Α	α	alpha
Β	β	beta
Γ	γ	gamma
Δ	δ	delta
Ε	ϵ	epsilon
Ζ	ζ	zeta
Η	η	eta
Θ	θ	theta
Ι	ι	iota
Κ	κ	kappa
Λ	λ	lambda
Μ	μ	mu
Ν	ν	nu
Ξ	ξ	xi
Ο	ο	omicron
Π	π	pi
Ρ	ρ	rho
Σ	σ	sigma
Τ	τ	tau
Υ	υ	upsilon
Φ	ϕ φ	phi
Χ	χ	chi
Ψ	ψ	psi
Ω	ω	omega

Characters not often used as symbols are shaded; for the most part they are too close in form to roman characters to be of value as independent symbols.

Contents

foreword, v
preface to the Berkeley Physics Course, vii
preface to volume I, ix
teaching notes, xi
note to the student, xv
notation, xvii

1. Introduction, 2
2. Vectors, 24
3. Galilean Invariance, 54
4. Simple Problems in Nonrelativistic Dynamics, 92
5. Conservation of Energy, 132
6. Conservation of Linear and Angular Momentum, 164
7. Harmonic Oscillator, 190
8. Elementary Dynamics of Rigid Bodies, 234
9. Inverse-Square-Law Force, 260
10. Speed of Light, 310
11. Lorentz Transformations of Length and Time, 344
12. Relativistic Dynamics: Momentum and Energy, 380
13. Simple Problems in Relativistic Dynamics, 406
14. Principle of Equivalence, 426
15. The Particles of Modern Physics, 438

index, 471
tables of values, inside covers

Chapter 1

The natural world 3

Geometry and physics 7
Estimates of the curvature of space 9
Planetary predictions 9
Trigonometrical parallax 10
Geometry on a smaller scale 11
Invariance 13

Further reading 14

Film list 14

Problems 14

Advanced Topic: Simple astronomy within the solar system 15

Experimental tools of physics 18

Introduction

The Natural World. To every man the natural world seems immense and complex, the stage for a startling diversity of appearances and events. These impressions are supported by estimates of the general order of magnitude of the values of some interesting quantities. At this stage we shall not enter into the arguments and measurements which lead to the figures given. The most remarkable thing about these numbers is that we know them at all; it is not of pressing importance that some of them are known only approximately.

The universe is immense. From astronomical observations we infer the value 10^{28} cm or 10^{10} light-years for a characteristic length which is loosely called the radius of the universe. The value is uncertain by perhaps a factor of 3. For comparison, the distance of the earth from the sun is 1.5×10^{13} cm and the radius of the earth is 6.4×10^{8} cm.

The number of atoms in the universe is very large. The total number of protons and neutrons in the universe, with an uncertainty perhaps of a factor of 100, is believed to be of the order of 10^{80}. Those in the sun number 1×10^{57} and those in the earth 4×10^{51}. The total in the universe would provide about $10^{80}/10^{57}$ (or 10^{23}) stars equal in mass to our sun. (This is one-sixth of a mole of stars.) Most of the mass of the universe is believed to lie in stars, and all known stars have masses between 0.01 and 100 times that of our sun.

Life is the most complex phenomenon in the universe. Man, one of the more complex forms of life, is composed of about 10^{16} cells. A cell is an elementary physiological unit that contains about 10^{12} to 10^{14} atoms. Every cell of every variety of living matter is believed to contain at least one long molecular strand of DNA (deoxyribonucleic acid). The DNA strands in a cell hold all the chemical instructions or genetic information needed to construct a complete man or bird or bacterium or tree. In a DNA molecule, which is composed of 10^8 to 10^{10} atoms, the precise arrangement of the atoms may vary from individual to individual; the arrangement always varies from species to species.† More than 10^6 species have been described and named on our planet.

Inanimate matter also appears in many forms. Protons, neutrons, and electrons combine to form about 100 different chemical elements and about 10^3 identified isotopes. The individual elements have been combined in various proportions to form perhaps 10^6 or more identified, differentiated chemical compounds, and to this number may be added a vast number of liquid and solid solutions and alloys of various compositions having distinctive physical properties.

Through experimental science we have been able to learn all these facts about the natural world, triumphing over darkness and ignorance to classify the stars and to estimate their masses, composition, distances, and velocities; to classify living species and to unravel their genetic relations; to synthesize inorganic crystals, biochemicals, and new chemical elements; to measure the spectral emission lines of atoms and molecules over a frequency range from 100 to 10^{20} cycles per second (cps); and to create new fundamental particles in the laboratory.

These great accomplishments of experimental science were achieved by men of many types: patient, persistent, intuitive, inventive, energetic, lazy, lucky, narrow, and with skilled hands. Some preferred to use only simple apparatus; others invented or constructed instruments of great refinement or size or complexity. Most of these men had in common only a few things: they were honest and actually made the observations they recorded, and they published the results of their work in a form permitting others to duplicate the experiment or observation.

The description we have given of the natural world as immense and complex is not the whole story, for theoretical understanding makes several parts of the world picture look much simpler. We have gained a remarkable understanding of some central and important aspects of the world. The areas which we believe we understand, summarized below, together with the theory of relativity and of statistical mechanics, are perhaps the greatest intellectual achievements of mankind.

---

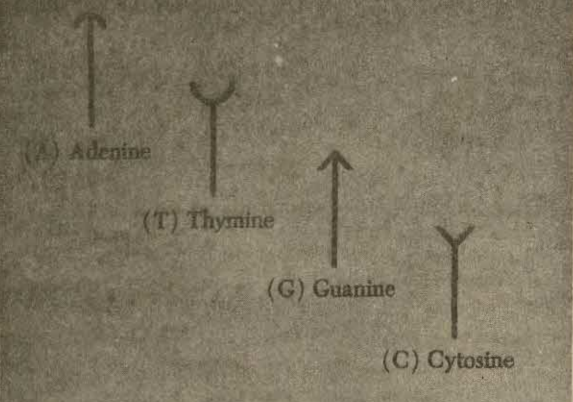
† The term *species* is defined roughly by the statement that two populations are different species if some describable difference(s) can be found between them and if they do not interbreed in a state of nature.

1. The laws of classical mechanics (Vol. I), which allow us to predict with remarkable accuracy the motions of the several parts of the solar system (including comets and asteroids), have led to the prediction and discovery of new planets. These laws suggest possible mechanisms for the formation of stars and galaxies, and together with the laws of radiation, they give a good account of the observed connection between the mass and luminosity of stars. The astronomical applications of the laws of classical mechanics are the most beautiful, but not the only successful, applications. We use the laws constantly in everyday life and in the engineering sciences.

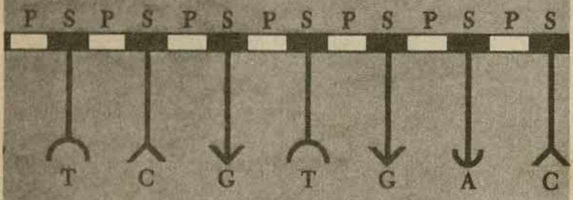
2. The laws of quantum mechanics (Vol. IV) give a very good account of atomic phenomena. For simple atoms predictions have been made which agree with experiment to 1 part in 10^5 or better. When applied to large-scale terrestrial and celestial events, the laws of quantum mechanics are to an excellent approximation identical with the laws of classical mechanics. Quantum mechanics provides, in principle, a precise theoretical basis for all of chemistry and metallurgy and for much of physics, but often we cannot handle the equations on existing or foreseeable computers. In some fields nearly all the problems seem too difficult for a direct theoretical attack based on first principles.

3. The laws of classical electrodynamics, which give an excellent account of all electric and magnetic effects except on the atomic scale, are the basis of the electrical engineering and communications industries. Electric and magnetic effects on an atomic scale are described exactly by the theory of quantum electrodynamics. Classical electrodynamics is the subject of Vols. II and III; some aspects of quantum electrodynamics are touched on in Vol. IV—a complete discussion must be deferred until a later course.

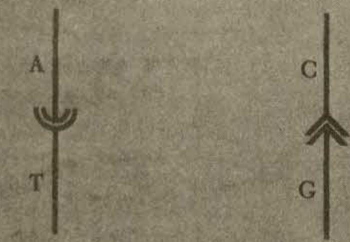
4. In a narrower example at another level, the operation of the genetic code now seems to be understood, and we find that the information storage of the cell of a simple organism exceeds that of the best present-day commercial computers. These matters are the subject of molecular biology: In nearly all life on our planet the complete coding of genetic information in the DNA molecule is carried by a linear sequence of four different molecular groups that are derived from the organic bases *adenine, thymine, guanine,* and *cytosine.* All genetic information in the cell is contained in



Schematic representation of the four nucleotide bases from which the DNA molecule is derived.



The nucleotides are connected to sugar groups S which, in turn, are bound to phosphate groups P to form a chain. The complete DNA molecule is composed . . .



of a double chain in the form of a helix. The two strands are connected by hydrogen bonds between adenine and thymine groups or between guanine and cytosine groups.

the order in which the nucleotide bases occur. Their chemical formulas do not concern us here, and we can call them A, T, G, and C, respectively. What is important is that in the biological replication of the DNA molecule an A on the old, or parent, strand can fit in comfortably only opposite a T on the new, or daughter, strand; similarly, a G can pair only with a C. If we write random combinations of the four letters A, T, G, C on a line, we have in the A-T and G-C pairing a specific, unambiguous instruction about how to write a second line:

TACGAACTTATCGCAA
ATGCTTGAATAGCGTT

The lines should be continued out to about 10^6 entries, and perhaps to 10^9 for the cell of a complex organism such as man.

The physical laws and theoretical understanding described in these examples are different in character from the direct results of experimental observations. The laws, which summarize the essential parts of a large number of observations, allow us to make successfully certain types of predictions, limited in practice by the complexity of the system. Often the laws suggest new and unusual types of experiments. Although the laws can usually be stated in compact form,† their application may sometimes require lengthy mathematical analysis and computation.

There is another aspect of the fundamental laws of physics: those laws of physics which we have come to understand have an attractive simplicity and beauty.‡ This does not mean that everyone should stop doing experiments, because the laws of physics have generally been discovered only after painstaking and ingenious experiment. The statement does mean that we shall be greatly surprised if future statements of physical theory should contain ugly and clumsy elements. The aesthetic quality of the discovered laws of physics colors our expectations about the laws still unknown. We tend to call a hypothesis attractive when its simplicity and elegance single it out among an infinite number of conceivable but incorrect theories.

---

† The first sentence of a short paperback is: "These lectures will cover all of physics." R. Feynman, *Theory of fundamental processes* (W. A. Benjamin, New York, 1961).

‡ "It seems that if one is working from the point of view of getting beauty in one's equations, and if one has really a sound insight, one is on a sure line of progress." P. A. M. Dirac, *Scientific American* **208** (5), 45–53 (1963). But most physicists feel the real world is too subtle for such bold attacks except by the greatest minds of the time, such as Einstein or Dirac or a dozen others. In the hands of a thousand others this approach has been limited by the inadequate distribution among men of "a sound insight."

In this course we shall make an effort to state some of the laws of physics from viewpoints which emphasize the features of simplicity and elegance. As we go along, we shall try also to give some of the flavor of good experimental physics, although this is very hard to do in a textbook. The research laboratory is the natural training ground for experimental physics.

Geometry and Physics

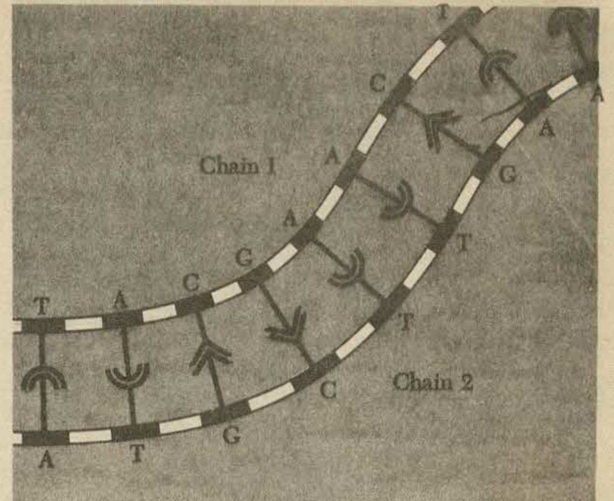
Mathematics, which permits the attractive simplicity and compactness of expression necessary for a reasonable discussion of the laws of physics and their consequences, is the language of physics. It is a language with special rules. If the rules are obeyed, only correct statements can be made: The square root of 2 is 1.414 . . . , or $\sin 2a = 2 \sin a \cos a$.

We must be careful not to confuse such truths with exact statements about the physical world. It is a question of experiment, rather than contemplation, to see whether the measured ratio of the circumference to the diameter of a physical circle really is 3.14159 . . . . Geometrical measurement is basic to physics and we must decide such questions before proceeding to use Euclidean or any other geometry in the description of nature. Here certainly is a question about the universe: Can we assume for physical measurements the truth of the axioms and theorems of Euclid?

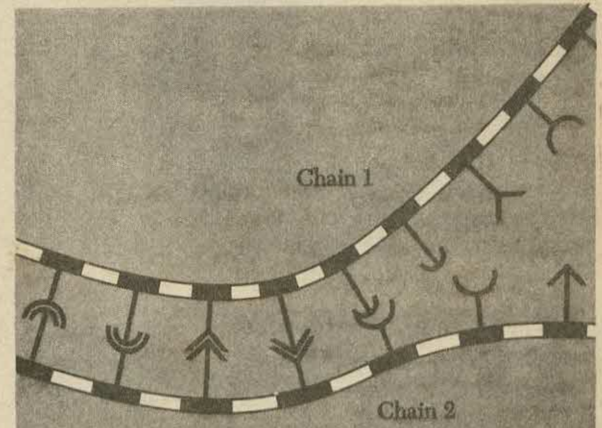
We can say only a few simple things about the experimental properties of space without becoming involved in difficult mathematics.

The most famous theorem in all mathematics is that attributed to Pythagoras: For a right-angled triangle the square of the hypotenuse equals the sum of the squares of the adjacent sides. Does this mathematical truth also hold in the physical world? Could it be otherwise? Contemplation of the question is insufficient and we must appeal to experiment for an answer. We give arguments which are somewhat incomplete because we are not able here to use the mathematics of curved three-dimensional space.

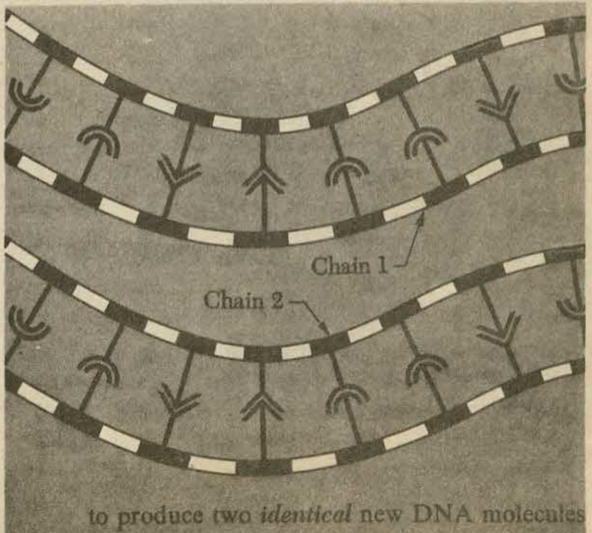
Consider first the situation of two-dimensional beings who live in a universe which is the surface of a sphere. Their mathematicians have described to them the properties of spaces of three or even higher dimensions, but they have as much difficulty in developing an intuitive feeling about such matters as we have in picturing a four-dimensional space. How can they determine whether they live on a curved surface? One way is to test the axioms of plane geometry by trying to confirm experimentally some of the theorems in Euclid. They will construct straight lines as the



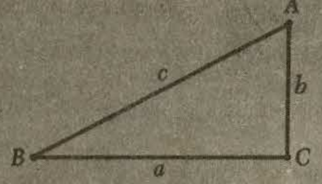
All genetic information in the cell is contained in the order in which the nucleotide bases occur.



When the cell reproduces, each DNA molecule splits into two separate chains. Each free chain then forms its complement from existing cell material . . .

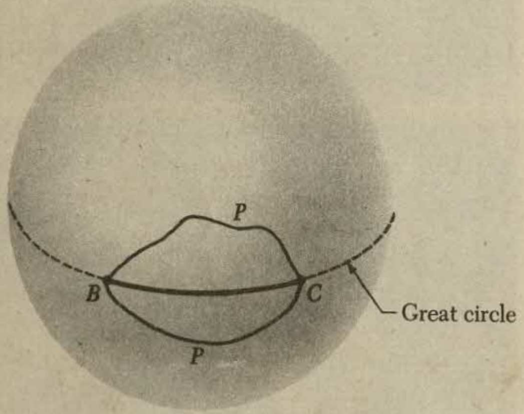


to produce two *identical* new DNA molecules

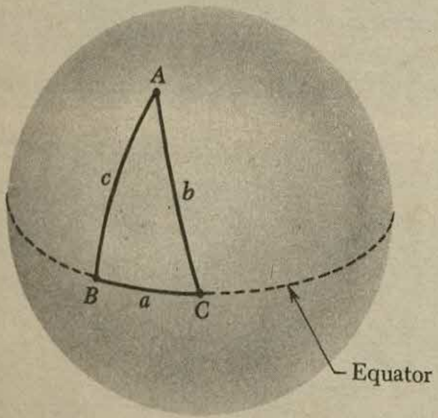


$$a^2+b^2=c^2$$

Do the axioms of Euclidean geometry, from which the Pythagorean theorem is logically derived, accurately describe the physical world? Only experiment can decide.



The shortest, "straight-line" distance between points *B* and *C* on a sphere lies along the great circle through these points and not along any other paths *P*.



Given three points *ABC*, the two-dimensional beings could construct triangles with "straight lines" as sides. They would find that for small right triangles $a^2 + b^2 \cong c^2$ and the sum of the angles of the triangle is slightly greater than 180°.

shortest path between any two points *B* and *C* on the surface of a sphere. We would describe such a path as a great circle.

They can go on to construct triangles and to test the Pythagorean theorem. For a very small triangle, each of whose sides is small in comparison with the radius of the sphere, the theorem would hold with great but not perfect accuracy; for a large triangle striking deviations would become apparent.

If *B* and *C* are points on the equator of the sphere, the "straight line" connecting them is the section of the equator from *B* to *C*. The shortest path from *C* on the equator to the north pole *A* is the line of fixed longitude which meets the equator *BC* at a right angle. The shortest path from *A* to *B* is a path of fixed longitude which also meets the equator *BC* at a right angle. Here we have a right triangle with $b = c$. The Pythagorean theorem is clearly invalid on the sphere because c^2 cannot now be equal to $b^2 + a^2$; further, the sum of the interior angles of the triangle *ABC* is always greater than 180°. Measurements made on the curved surface by its two-dimensional inhabitants enable them to demonstrate for themselves that the surface is indeed curved.

It is always possible for the inhabitants to say that the laws of plane geometry adequately describe their world but the trouble lies with the meter sticks used to measure the shortest path and thus define the straight line. The inhabitants could say that the meter sticks are not constant in length but stretch and shrink as they are moved to different places on the surface. Only when it is determined by continued measurements in different ways that the same results always hold does it become evident that the simplest description of why Euclidean geometry fails lies in the curvature of the surface.

The axioms of plane geometry are not self-evident truths in this curved two-dimensional world; they are not truths at all. We see that the actual geometry of the universe is a branch of physics which must be explored by experiment. We do not customarily question the validity of Euclidean geometry to describe measurements made in our own three-dimensional world because Euclidean geometry is such a good approximation to the geometry of the universe that any deviations from it do not show up in practical measurements. This does not mean that the applicability of Euclidean geometry is self-evident or even exact. It was suggested by the great nineteenth-century mathematician Carl Friedrich Gauss that the Euclidean flatness of three-dimensional space should be tested by measuring the sum of the interior angles of a large triangle; he realized that if three-dimensional space is curved, the sum of the angles of a *large enough* triangle might be significantly different from 180°.

Gauss† used surveying equipment (1821–1823) to measure accurately the triangle in Germany formed by Brocken, Hohehagen, and Inselberg. The longest side of the triangle was about 100 km. The measured interior angles were

$$\begin{array}{rr} & 86°13'58.366'' \\ & 53°\ \ 6'45.642'' \\ & 40°39'30.165'' \\ \hline \text{Sum} & 180°00'14.173'' \end{array}$$

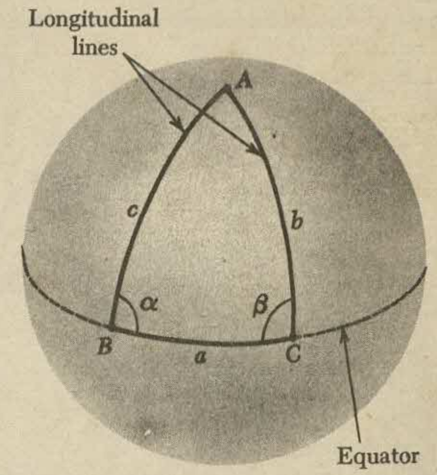
(We have not found a statement about the estimated accuracy of these values; it is likely that the last two decimal places are not significant.) Because the surveying instruments were set up *locally* horizontal at all three vertices, the three horizontal planes were not parallel. A calculated correction called the *spherical excess,* which amounts to 14.853 seconds of arc, must be subtracted from the sum of the angles. The sum thus corrected,

$$179°59'59.320'',$$

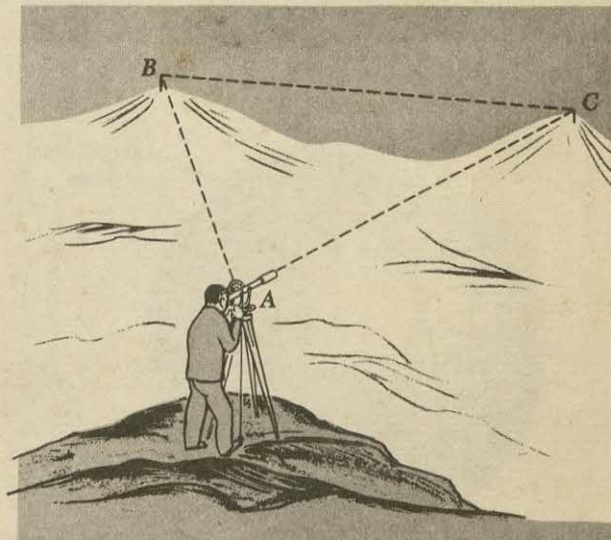
differs by 0.680 seconds of arc from 180°. Gauss believed this to lie within the observational error, and he concluded that space was Euclidean within the accuracy of these observations.

We saw in the earlier example that Euclidean geometry adequately described a small triangle on the two-dimensional sphere but departures became more evident as the scale increased. To see if our own space is indeed flat we need to measure very large triangles whose vertices are formed by the earth and distant stars or even galaxies. But we are faced with a problem. Our position is fixed by that of the earth and we are not yet free to wander through space with meter sticks to measure astronomical triangles. How can we test the validity of Euclidean geometry to describe measurements in space?

Estimates of the Curvature of Space. *Planetary Predictions.* A first lower limit of about 5×10^{17} cm to the radius of curvature of our own universe is implied by the consistency of astronomical observations within the solar system. For example, the positions



If they used larger triangles, the sum of the angles would become increasingly greater than 180°. Here, with B and C on the equator, and A on the pole, α and β are *both* right angles. Obviously $a^2 + b^2 \neq c^2$, because b is equal to c.



Gauss measured the angles of a triangle with vertices on three mountain tops and found no deviation from 180°, within the accuracy of his measurements.

---

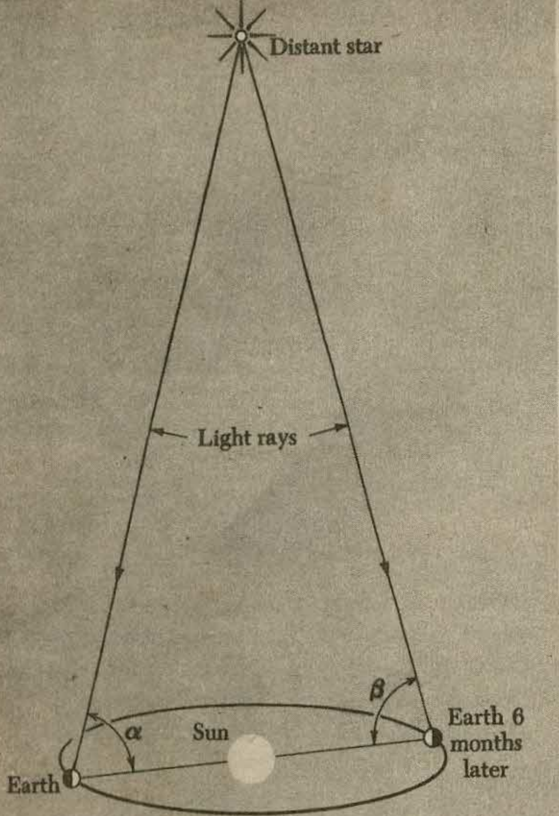
† C. F. Gauss, *Werke,* vol. 9, see especially pp. 299, 300, 314, and 319. The collected works of Gauss are a remarkable example of how much a gifted man can accomplish in a lifetime.

of the planets Neptune and Pluto were inferred by calculation before their visual confirmation by telescopic observation. Small perturbations of the orbits of the known planets led to the discovery of Neptune and Pluto very close to the positions calculated for them. We can easily believe that a slight error in the laws of geometry would have destroyed this coincidence. The outermost planet in the solar system is Pluto. The average radius of the orbit of Pluto is 6×10^{14} cm; the closeness of the coincidence between the predicted and observed positions implies a radius of curvature of space of at least 5×10^{17} cm. An infinite radius of curvature (flat space) is not incompatible with the data. It would take us too far from our present purpose to discuss the numerical details of how the estimate of 5×10^{17} cm is arrived at or to define precisely what is meant by the radius of curvature of a three-dimensional space. The two-dimensional analog of the surface of a sphere can be used in this emergency as a useful crutch.

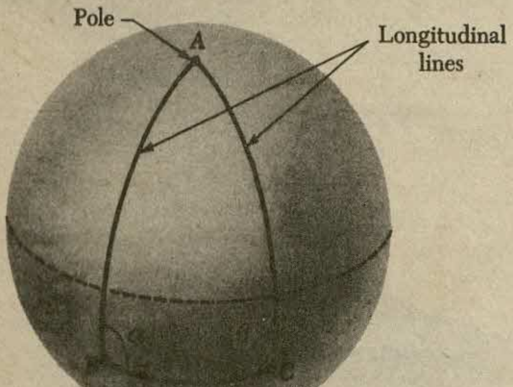
Trigonometrical Parallax. Another argument was suggested by Schwarzschild.† In two observations taken 6 months apart, the position of the earth relative to the sun has changed by 3×10^{13} cm, the diameter of the earth's orbit. Suppose that at these two times we observe a star and measure the angles α and β. Here α and β are the Greek characters alpha and beta. If space is flat, the sum of the angles $\alpha + \beta$ is always less than 180°, and the sum approaches this value as the star becomes infinitely distant. One-half of the deviation of $\alpha + \beta$ from 180° is called the *parallax*. But in a curved space it is not necessarily true that $\alpha + \beta$ is always less than 180°.

We return to our two-dimensional astronomers living on the surface of a sphere to see how they discover that their space is curved from a measurement of the sum $\alpha + \beta$. From our previous discussion of the triangle *ABC* we see that, when the star is a quarter of a circumference away, $\alpha + \beta = 180°$. When the star is nearer, $\alpha + \beta < 180°$; when it is farther away, $\alpha + \beta > 180°$. The astronomer need merely look at stars more and more distant and measure $\alpha + \beta$ to see when the sum begins to exceed 180°. The same argument is valid within our three-dimensional space.

There is no observational evidence that $\alpha + \beta$ as measured by astronomers is ever greater than 180°, after an appropriate correc-



Schwarzschild's demonstration that on a flat surface $\alpha + \beta < 180°$. The *parallax* of a star is defined as $\frac{1}{2}(180° - \alpha - \beta)$.



For this triangle, with *B* and *C* below the equator, $\alpha + \beta > 180°$, which can only happen because the two-dimensional "space" of the spherical surface is curved. A similar argument can be applied to three-dimensional space. The radius of curvature of the two-dimensional space shown here is just the radius of the sphere.

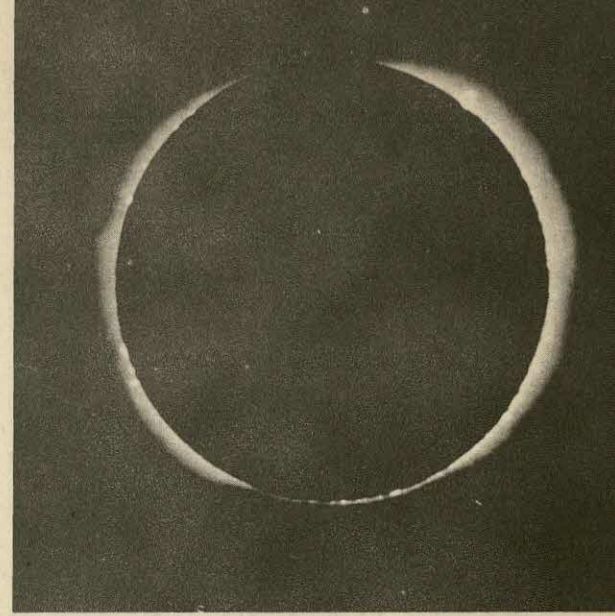
---

† K. Schwarzschild, *Vierteljahrsschrift der astronomischen Gesellschaft* 35, 337 (1900).

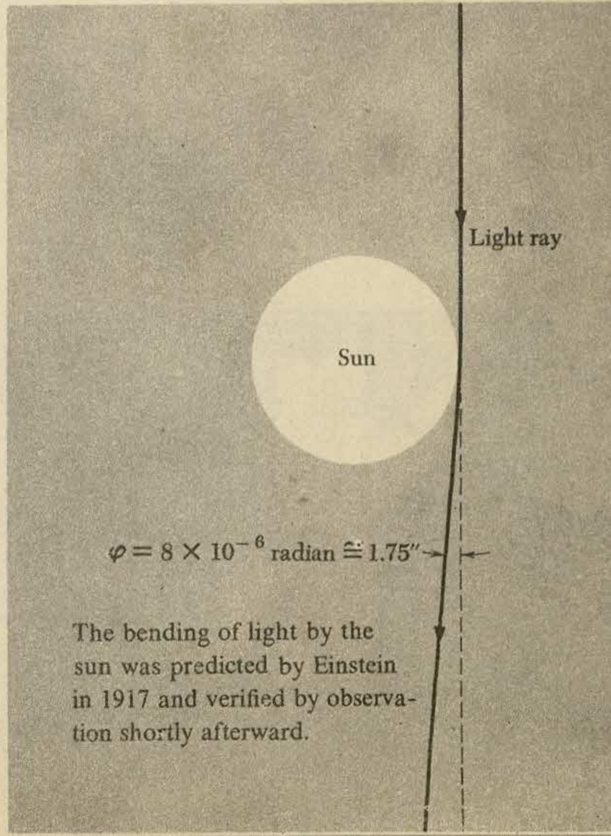
tion is made for the motion of the sun relative to the center of our galaxy. Values of $\alpha + \beta$ less than 180° are used to determine by triangulation the distances of nearby stars. Values less than 180° can be observed out to about 3×10^{20} cm,† the limit of angle measurement with present telescopes. It cannot be inferred directly from this argument that the radius of curvature of space must be larger than 3×10^{20} cm; for some types of curved space other arguments are needed. The answers come out finally that the radius of curvature (as determined by triangulation) must be larger than 6×10^{19} cm.

At the beginning of Chap. 1 we said that a characteristic length associated with the universe, which has a value of the order of 10^{28} cm or 10^{10} light-years, is observed. The most simple-minded interpretation of this length calls it the radius of the universe. Another possible interpretation gives it as the radius of curvature of space. Which is it? This is a cosmological question, an excellent introduction to the speculative science of cosmology is given in the book by Bondi cited at the end of this chapter. We summarize our knowledge of the radius of curvature of space by the statements that it is not smaller than 10^{28} cm and that we do not know that space on a large scale is not flat.

Geometry on a Smaller Scale. The foregoing observations bear upon the average radius of curvature of space and are not sensitive to bumps which are believed to exist in the immediate neighborhood of individual stars and which contribute a local roughness to the otherwise flat, or slightly curved, space. Experimental data which bear upon this question are extremely hard to acquire, even for the neighborhood of our sun. By careful and difficult observations of stars visible near the edge of the sun during a solar eclipse it has been established that light rays are slightly curved when they pass near the edge of the sun and, by inference, close to any similarly massive star. For a grazing ray the angle of bend is very slight, amounting to only 1.75″. Thus as the sun moves through the sky the stars which are almost eclipsed would, if we could see them in the daytime, appear to spread out very slightly from their normal positions. This merely says that the light moves in a curved path near the sun; it does not by itself insist upon the unique interpretation that the space around the



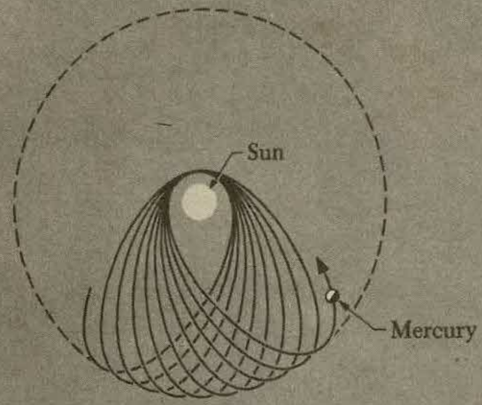
Solar eclipse of July 20, 1963. (*Photographed by C. H. Cleminshaw, Griffith Observatory.*)



The bending of light by the sun was predicted by Einstein in 1917 and verified by observation shortly afterward.

---

† It may be objected that the distance measurements themselves assume that Euclidean geometry is applicable. Other methods of estimating distance are available, however, which are discussed in modern texts on astronomy.

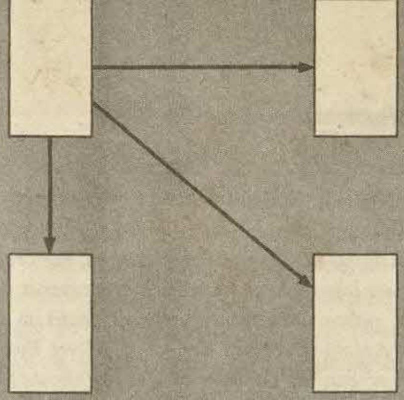


Precession of the orbit of Mercury explained by the general theory of relativity. The plane of the orbit is in the page; the eccentricity of the orbit is greatly exaggerated for clarity. Without precession the figure would be a stationary ellipse.

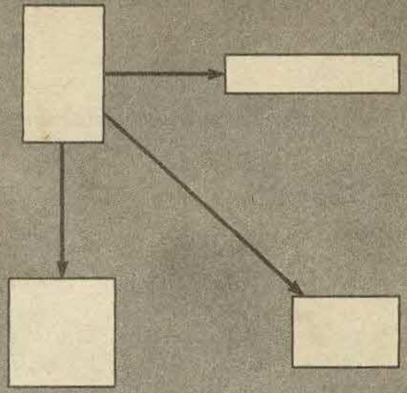
sun is curved. Only with accurate measurements by meter sticks of various materials close to the sun's surface could we establish directly that a curved space is the most efficient and natural description. One other kind of observation bears upon the possibility of a curved space. The orbit of Mercury, the planet nearest the sun, differs very slightly from that predicted by application of Newton's laws of universal gravitation and motion, even after certain small corrections of the special theory of relativity have been incorporated into the calculated orbit. Could this be an effect of curved space near the sun? To answer such a question we should have to know how a possible curvature would affect the equations of motion for Mercury, and this involves more than just geometry.

In a remarkable and beautiful series of papers, Einstein (1917) described a theory of gravitation and geometry, the general theory of relativity, which predicted, in quantitative agreement with the observations, just the two effects described above. These are still the only crucial confirmations of the geometric predictions of the theory. Despite the meager evidence the essential simplicity of the general theory has made it widely accepted.

From astronomical measurements we concluded that Euclidean geometry gives an extraordinarily good description of measurements of lengths, areas, and angles, at least until we reach the enormous lengths of the order of 10^{28} cm. But so far nothing has been said about the use of Euclidean geometry to describe very small configurations comparable in size to the 10^{-8} cm of an atom or the 10^{-12} cm of a nucleus. The question of the validity of Euclidean geometry ultimately must be phrased as follows: Can we make sense of the subatomic world, can we make a successful physical theory to describe it, while assuming that Euclidean geometry is valid? If we can, then there is no reason at present to question Euclidean geometry as a successful approximation. We shall see in Vol. IV that the theory of atomic and subatomic phenomena does not seem to lead to any paradoxes which have thus far blocked our understanding of them. Many facts are not understood, but none appear to lead to contradictions. In this sense Euclidean geometry stands the test of experiment down at least to 10^{-13} cm.



Invariance under translation. Moving object to any other place does not alter its size or shape.



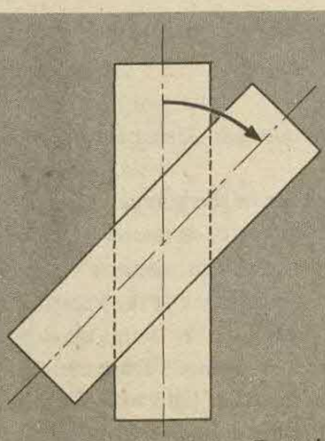
Noninvariance under translation in a hypothetical world. Moving object to another place may alter its size or shape.

Invariance. We may summarize some of the consequences of the experimental validity of Euclidean geometry:

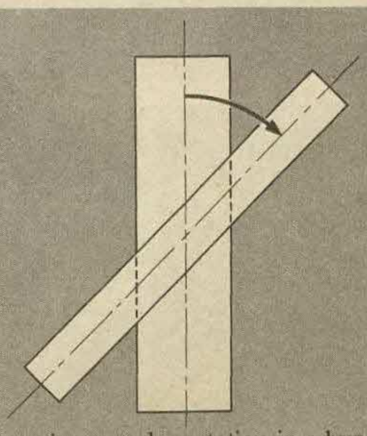
Invariance under translation. By this we mean that our space is homogeneous, that it is no different from point to point. If figures are moved without rotation, there is no change in their properties.

Invariance under rotation. By experiment it is known that space is isotropic to high precision, so that all directions are equivalent; figures are unaltered by rotation. It is possible to imagine a flat space which is not isotropic. For example, the speed of light in some direction could be twice what it is in another direction at right angles to the first. But there is no evidence in free space for an effect of this kind. Within a crystal, however, many anisotropic effects are encountered.

The property of invariance under translation leads to the conservation of linear momentum; invariance under rotation leads to the conservation of angular momentum. These subjects are developed in Chaps. 3 and 6. The concept of invariance is developed in Chap. 2 and at the end of Chap. 3.



Invariance under rotation. Rotation of object does not alter its size or shape.



Noninvariance under rotation in a hypothetical world. Rotation of object may change size or shape.

Further reading

Physical Science Study Committee (PSSC), *Physics*, chaps. 1–4 (D. C. Heath and Company, Boston, 1960).

R. H. Baker, *Astronomy*, 7th ed. (Van Nostrand, Princeton, N.J., 1959).

H. Bondi, *Cosmology*, 2d ed. (Cambridge University Press, New York, 1960). Brief, clear, authoritative account, with emphasis on the observational evidence.

A. Einstein, "Autobiographical notes," in *Albert Einstein: philosopher-scientist*, edited by P. A. Schilpp (Library of Living Philosophers, Evanston, 1949). An excellent short autobiography. It is a pity that there are so few really great biographies of outstanding scientists, such as that of Freud by Ernest Jones. There is little else comparable in depth and in honesty to the great literary biographies, such as *James Joyce*, by Richard Ellman. The autobiography of Charles Darwin is a remarkable exception. Writers about scientists appear to be overly intimidated by Einstein's sentence: "For the essential of a man like myself lies precisely in *what* he thinks and *how* he thinks, not in what he does or suffers."

Larousse encyclopedia of astronomy (Prometheus Press, New York, 1962). This is a beautiful and informative book.

D. J. de Solla Price, *Little science, big science* (Columbia University Press, New York, 1963). A statistical and sociological study of the science of science.

Ann Roe, *The making of a scientist* (Dodd Mead and Co., New York, 1953; Apollo reprint, 1961). This is an excellent sociological study of a group of leading American scientists of the late 1940s. There have probably been some significant changes in the scientific population since the book was first published in 1953.

O. Struve, B. Lynds, and H. Pillans, *Elementary astronomy* (Oxford University Press, New York, 1959). Emphasizes the main ideas of physics in relation to the universe; excellent.

Film list

"Measuring Large Distances" (29 min) F. Watson (PSSC-MLA 0103). Shows how by triangulation and parallax measurements, distances to the moon and then to stars up to 500 light-years away can be measured.

"Change of Scale" (23 min) R. W. Williams (PSSC-MLA 0106). For orientation in the ideas of estimating and scaling. Gives several nice examples of scaling stresses and of scaling where something depends on the scale of the velocity (ship design).

Problems

1. *The known universe.* Using information in the text, estimate the following:

(*a*) The total mass in the known universe. *Ans.* $\sim 10^{56}$ gm.

(*b*) The average density of matter in the universe.

Ans. $\sim 10^{-29}$ gm/cm^3, equivalent to 100 hydrogen atoms/cm^3.

(*c*) The ratio of the radius of the known universe to that of the proton. (Take radius of proton as 10^{-13} cm; mass of proton $= 1.7 \times 10^{-24}$ gm.)

2. *Signals across a proton.* Estimate the time required for a signal traveling with the speed of light to move a distance equal to the diameter of a proton. Take the diameter of the proton to be 2×10^{-13} cm.

3. *Distance of Sirius.* The parallax of a star is one-half the angle subtended at the star by the extreme points in the earth's orbit around the sun. The parallax of Sirius is 0.371″. Find its distance in centimeters, light-years, and parsecs. (See the table of values inside the cover.)

Ans. 8.3×10^{18} cm; 8.8 light-years; 2.7 parsecs.

4. *Size of atoms.* Using the value of Avogadro's number given in the table and your estimate of an average density for common solids, estimate roughly the diameter of an average atom.

5. *Angle subtended by moon.* Obtain a millimeter scale and, when viewing conditions are favorable, try the following experiment: Hold the scale at arm's length and measure the diameter of the moon. Measure the distance from the scale to your eye. (The radius of the moon's orbit is 3.8×10^{10} cm, and the radius of the moon itself is 1.7×10^{8} cm.)

(*a*) If you were able to try the measurement, what was the result?

(*b*) If the measurement could not be made, from the data given above calculate the angle subtended by the moon at the earth.

Ans. 9×10^{-3} radian.

(*c*) What is the angle subtended at the moon by the earth?

Ans. 3.4×10^{-2} radian.

6. *Compton wavelength.* Construct an expression for a quantity having the dimensions of a length, starting from the velocity of light c, the mass of the electron m, and Planck's constant h. The dimensions of Planck's constant are [energy-time] or [mass] [length^2/time^2] [time]. Evaluate the length from the values of c, m, and h given in the table of values. This length plays an important role in atomic physics. It may be denoted by λ_C, where λ is the Greek character lambda.

Advanced Topic. Simple astronomy within the solar system

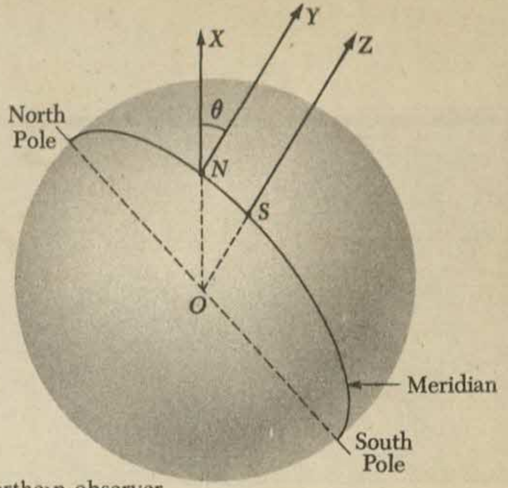
Two amateur astronomers set for themselves the goal of determining the diameter and mass of the sun. After giving the problem considerable thought, they realize that, to begin with, several subsidiary quantities must be established.

Their first step is to determine the radius of the earth, which they accomplish by using accurate maps to locate themselves 500 miles apart on a geographical meridian (line of longitude). They communicate by short-wave radio. The southern observer S selects a star which passes through his zenith at some instant of time.

At the time the selected star is passing through the zenith of the southern observer, it is also crossing the meridian of the northern observer N, but as a consequence of the earth's curvature, it crosses below his zenith.

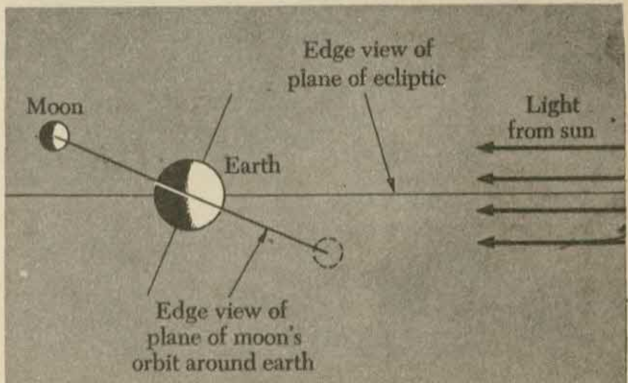
(*a*) If the northern observer notes a zenith angle of $\theta = 7.2°$, show that the radius of the earth is calculated to be 6.4×10^{8} cm.

The two amateurs find that the second step necessary is to measure the velocity of the moon in its orbit about the center of the earth. They accomplish this indirectly by determining the times at which some par-

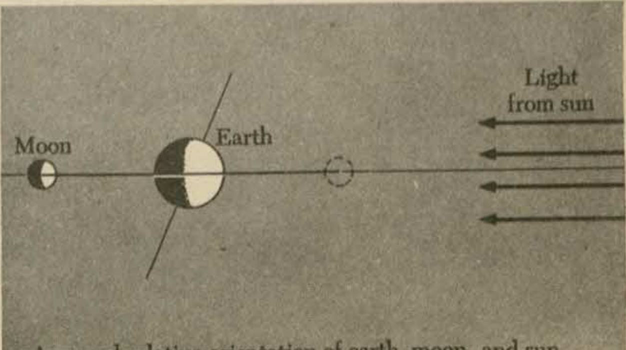


N = Northern observer
S = Southern observer
O = Center of earth

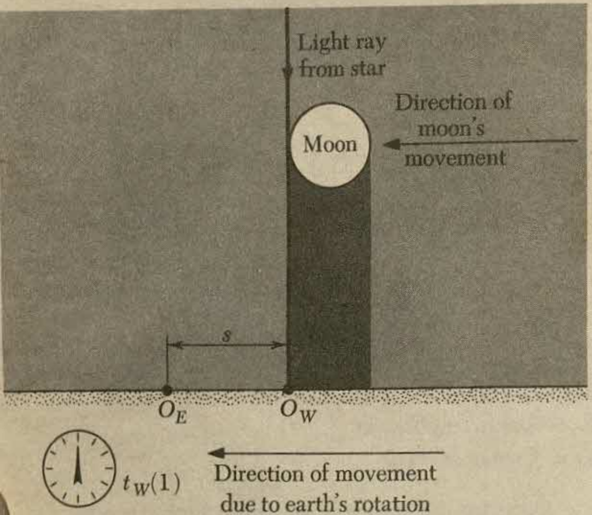
Light from distant star at zenith is received by S along line ZSO (line from zenith for S).
Light from same star is received by N along line $YN \parallel ZSO$.
The line XNO is directed from the zenith for N.
YN and XNO are inclined at zenith angle θ.


Edge view of plane of ecliptic
Moon
Light from sun
Earth
Edge view of plane of moon's orbit around earth

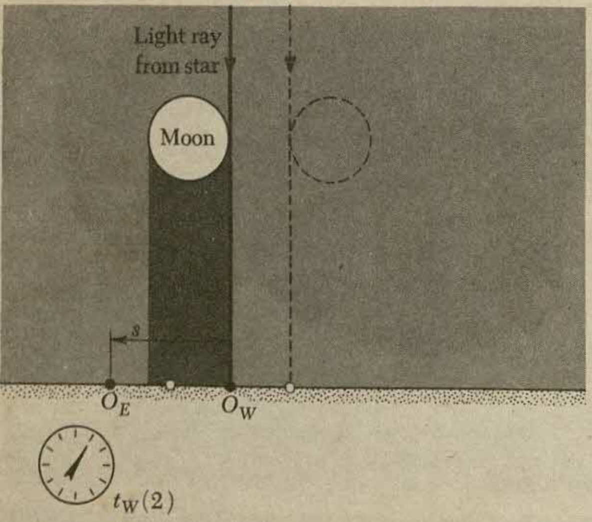
Actual relative orientation of earth, moon, and sun.



Assumed relative orientation of earth, moon, and sun (for purposes of discussion).


Light ray from star
Direction of moon's movement
Moon
s
O_E
O_W
$t_W(1)$
Direction of movement due to earth's rotation

Position of moon and observers O_W and O_E at time $t_W(1)$. Starlight is assumed to come from an infinite distance and therefore has parallel rays.



At time $t_W(2)$ the moon has moved to this position, and starlight now reappears to observer O_W who, because of the earth's rotation, has moved in the meantime to his new position shown here.

ticular star, as seen from two different points on the earth's surface, is eclipsed by the moon. To reduce the problems of geometrical and mathematical calculation they make a number of simplifying assumptions: the moon and the star are located in the ecliptic (the plane of the sun and of the earth's orbit); the star passes directly behind the moon, i.e., along a diameter; the observations are made at midnight with a full moon. The curvature of the earth, effects of atmospheric refraction, and further corrections are neglected. The geometrical situation is pictured on page 15.

Parallel rays of light from a distant star are received by the two observers O_W and O_E. The westerly observer O_W notes the time $t_W(1)$ at which the star disappears behind the moon and later the time $t_W(2)$ at which it reappears. Similar observations $t_E(1)$ and $t_E(2)$ are made by the easterly observer.

(*b*) Show that the velocity of the moon relative to the center of the earth is given by the expression

$$v_M = v_O + \frac{S}{t_E(1) - t_W(1)}$$

where S is the separation of the observers, v_O the linear velocity of the observers, v_M the linear velocity of the moon, and $t_E(1)$ and $t_W(1)$ are as defined above.

Knowing the velocity of the moon, either observer will be able to determine the moon's diameter from the total time that the moon eclipses the star.

(*c*) Show that the moon's diameter is

$$2R_M = (v_M - v_O)(t_2 - t_1)$$

where t_1 and t_2 are the set of times measured by either observer.

(*d*) From the radius of the earth determined above and the earth's known period of rotation determine its surface velocity at 30°N latitude relative to its center. (*Ans.* 4.04×10^4 cm/sec)

Typical observations which might have been obtained by the two amateurs are given in the following table.

Observation	$t_W(1)$	$t_W(2)$	$t_E(1)$	$t_E(2)$
Time, min	0.0	95.6	22.0	117.7

(*e*) Use these data and previous results to show that the velocity of the moon is 10.1×10^4 cm/sec and its diameter is 3.48×10^8 cm. Having determined the moon's orbital velocity, the two amateurs quickly use its known orbital period (2.36×10^6 sec) to determine the radius of its orbit.

(*f*) Show that their observations lead to a result in agreement with accurate measurements (3.8×10^{10} cm).

8733
Govt.
of Education

If the orbit of a satellite is circular, it is quite simple to determine the mass of the attracting body. Using Newton's law of gravitation $F = GM_EM_M/r^2$ for the force between the earth and the moon we show in Chap. 3 that

$$GM_E = v_M^2r = R^2g$$

where G is the gravitational constant, M_E the mass of the earth, v_M the velocity of the moon, r the radius of the moon's orbit, R the radius of the earth, and g the acceleration of gravity at the earth's surface (980 cm/sec^2). The first equality above is obtained by equating the gravitational force to the centrifugal force $M_Mv_M^2/r$, where M_M is the mass of the moon.

(*g*) Calculate the magnitude of the constant GM_E.

After considerable effort the amateurs conclude that, because they do not have sufficient information to calculate the earth's mass, they must perform another experiment. Ideally, they should measure the gravitational constant G. Because this is a fairly difficult experiment to carry out, they decide instead to estimate the density of the earth. From a study of surface materials they obtain a value of 5 gm/cm^3.

(*h*) What is the approximate mass of the earth on this basis? To what percentage error does this actually correspond?

(*i*) Use this estimate and preceding results to obtain the order of magnitude of G.

Next the two amateurs determine the distance to the sun by using the known data about the moon's orbit as shown in the illustration. At a particular time the sun, moon, and earth are in positions such that the shadow edge coincides with a major diameter of the moon (which to an observer appears to be exactly a half moon). At this time the observer determines the angle α between lines to the moon and to the sun.

(*j*) For an angle $\alpha = 89°51'$, calculate the distance r from the earth to the sun.

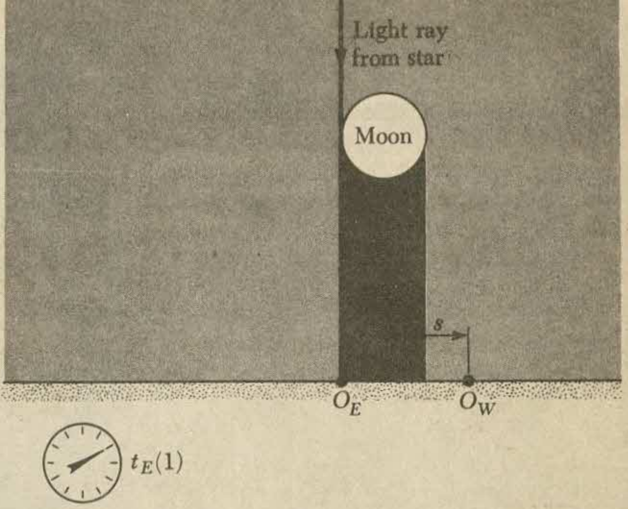
Knowing the distance of the earth from the sun, the two amateurs realize that Kepler's third law (in addition to Newton's law of motion) will allow them to calculate the mass of the sun (see PSSC, pp. 354–357).

(*k*) Calculate the mass of the sun.

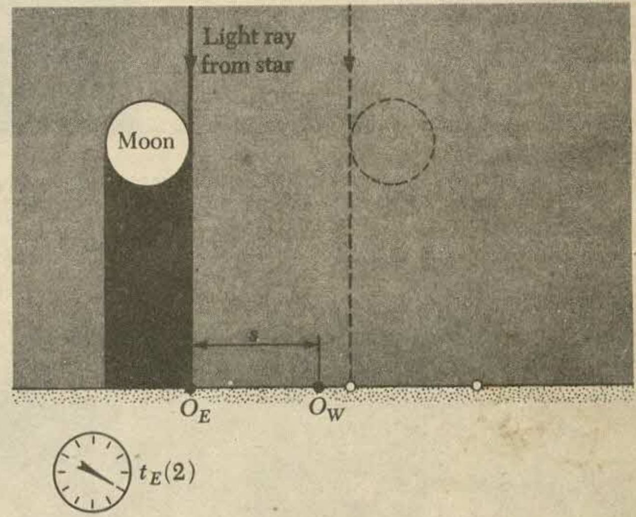
Finally the two amateurs, who had been fortunate enough to view the 1963 total eclipse of the sun, remember that the moon almost exactly obscured the sun. Calculate the radius R_S of the sun.

(*l*) Determine the diameter of the sun on the basis of information obtained above.

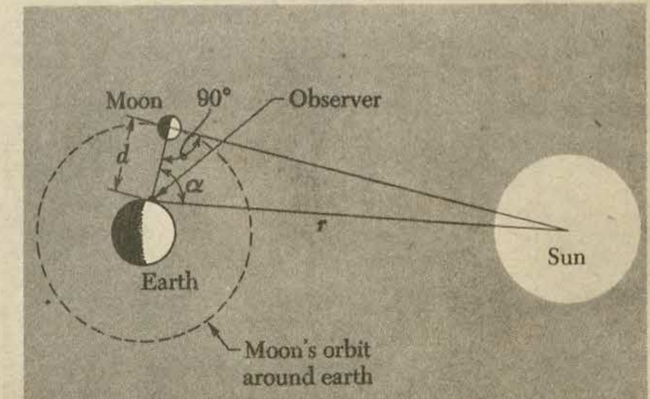
To make determinations such as these with extreme accuracy requires a great deal of work in instrumentation, observation, interpretation, calculation, and theory—matters which have occupied the minds of many men over the centuries and are of continuing interest. Considerations of accuracy, however, have not prevented men from making measurements in new areas. Were physics to wait for instruments of ultimate precision to be perfected, it might not advance at all. Many elaborate experimental installations, it has been said, are monuments to procrastination.



In the same way, O_E sees star disappear at $t_E(1)$ and after an interval . . .



reappear to him at time $t_E(2)$.



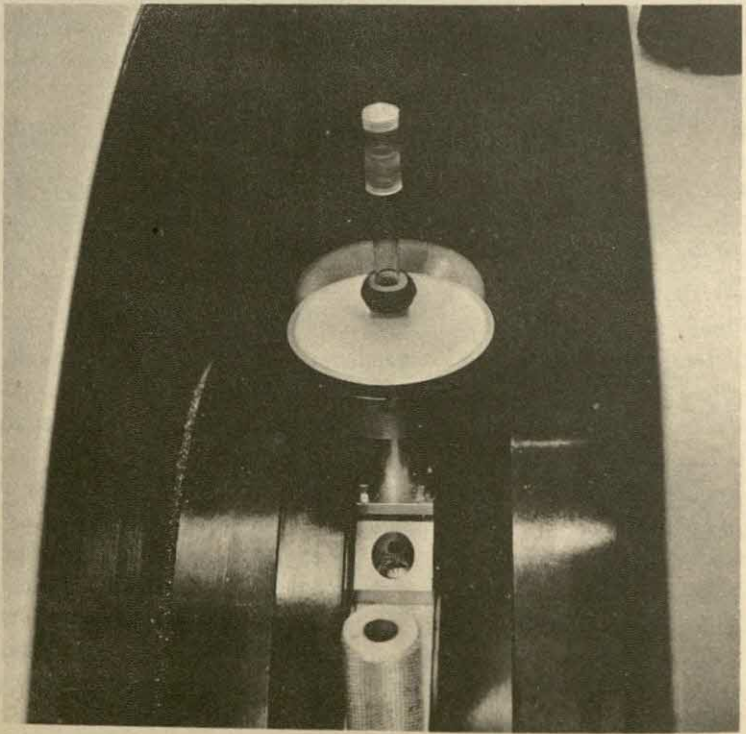
Method of determining distance from earth to sun using known data about the moon's orbit.

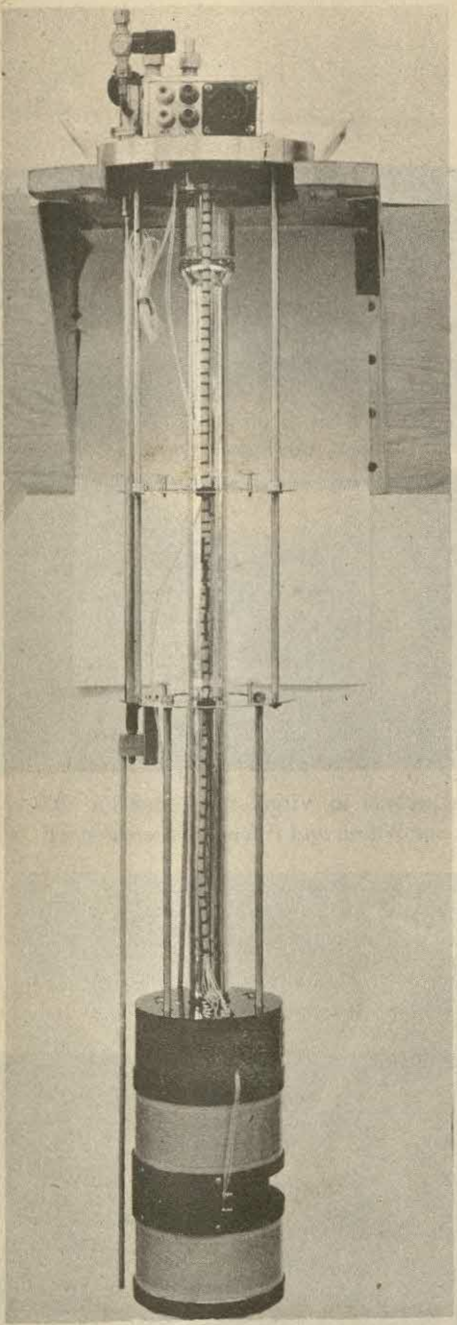
Experimental Tools of Physics
The photographs on these pages show some of the instruments and machines which are contributing actively to the advancement of the physical sciences.

A nuclear magnetic resonance laboratory for chemical structure studies. (*ASUC Photography*)

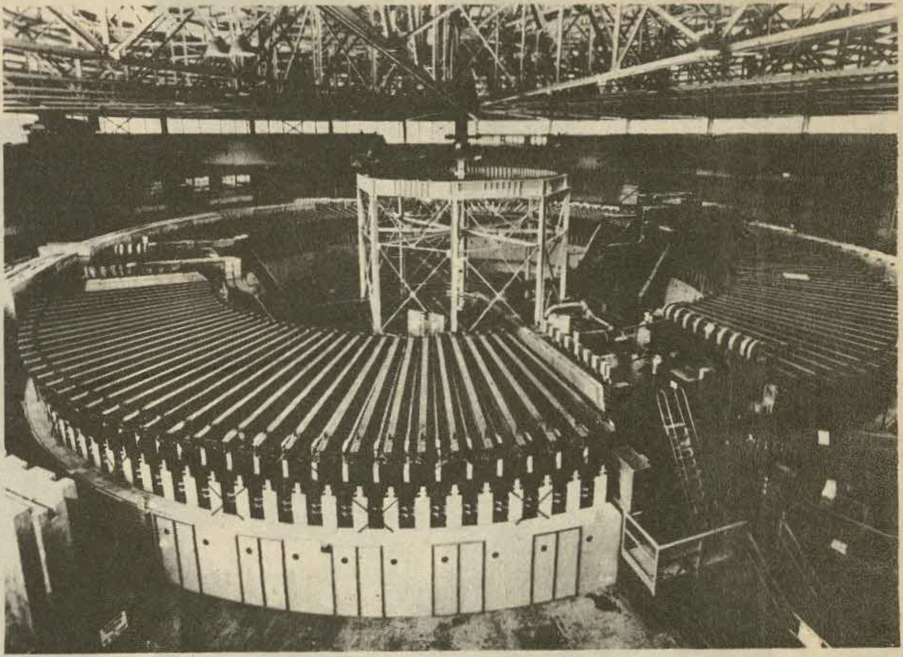
Operator in a nuclear magnetic resonance laboratory ready to place a sample in the probe in the variable temperature controller in which the sample is spun. (*Esso Research*)

Study of nuclear magnetic resonance spectra: a sample is shown spinning rapidly between the polepieces of an electromagnet to average out magnetic field variations. (*Esso Research*)



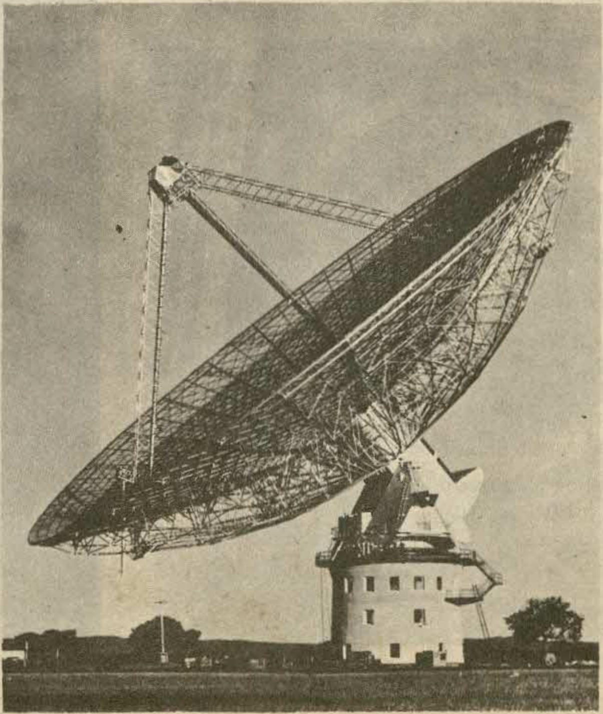


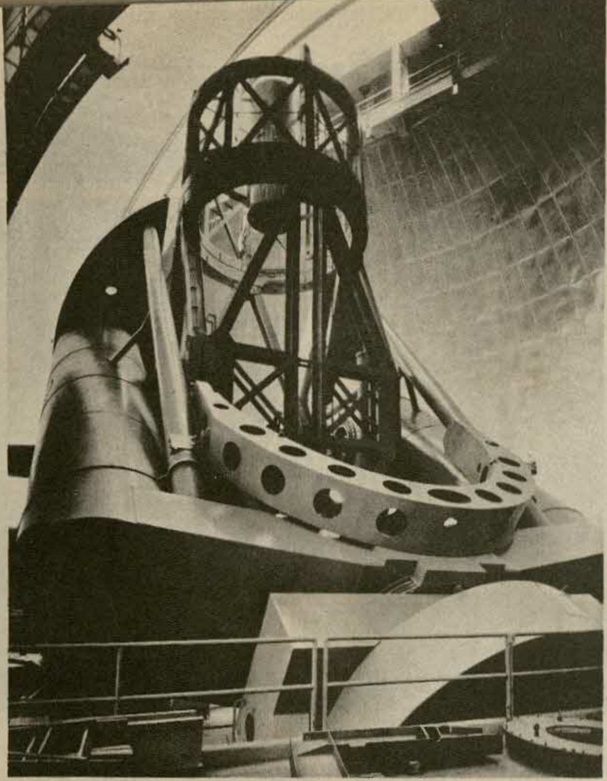
A magnet constructed of superconducting wire, for operation at low temperature. The coils shown are rated to produce a magnetic field of 54,000 gauss. Such apparatus is the heart of a modern low-temperature laboratory. (*Varian Associates*)



A high-energy particle accelerator: the Bevatron at Berkeley. Protons are injected at the lower right. (*Lawrence Radiation Laboratory*)

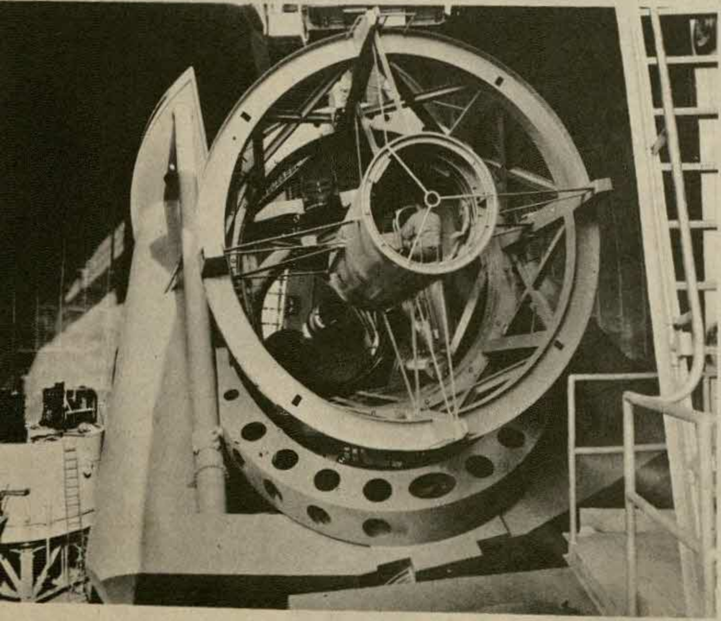
The large radio telescope in Australia. The dish is 210 ft. in diameter. It stands in a quiet valley 200 miles west of Sydney, New South Wales. In this remote location, there is a minimum of electrical interference. (*Australian News and Information Bureau*)

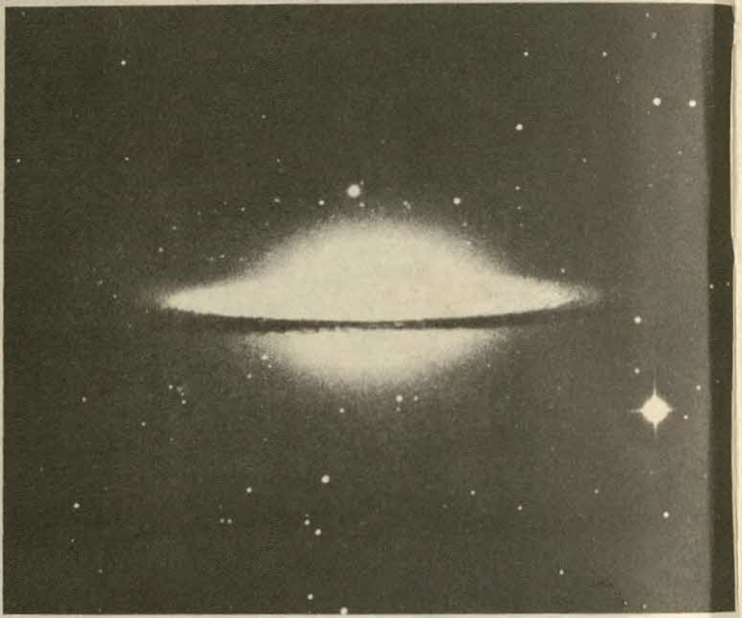




The 200-in. Hale telescope pointing to zenith; seen from the south. (*Mount Wilson and Palomar Observatories*)

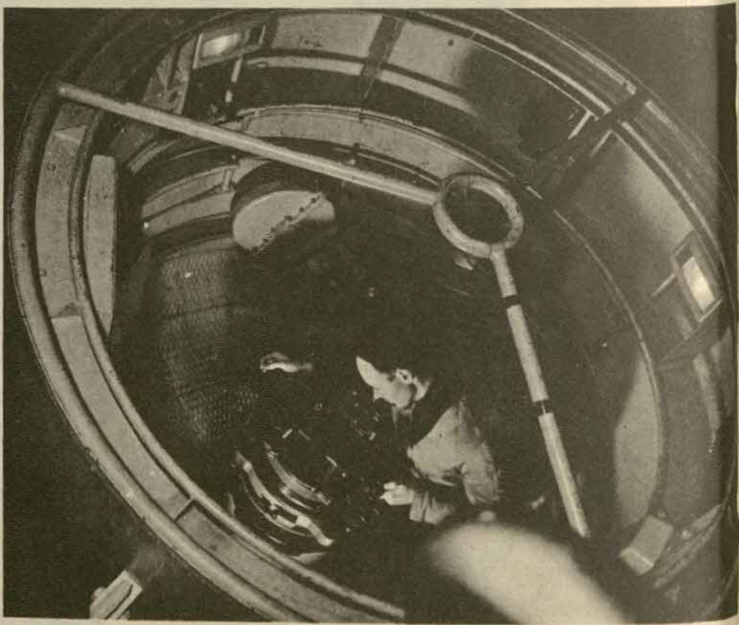
Reflecting surface of 200-in. mirror of Hale telescope and observer shown in prime-focus cage. (*Mount Wilson and Palomar Observatories*)

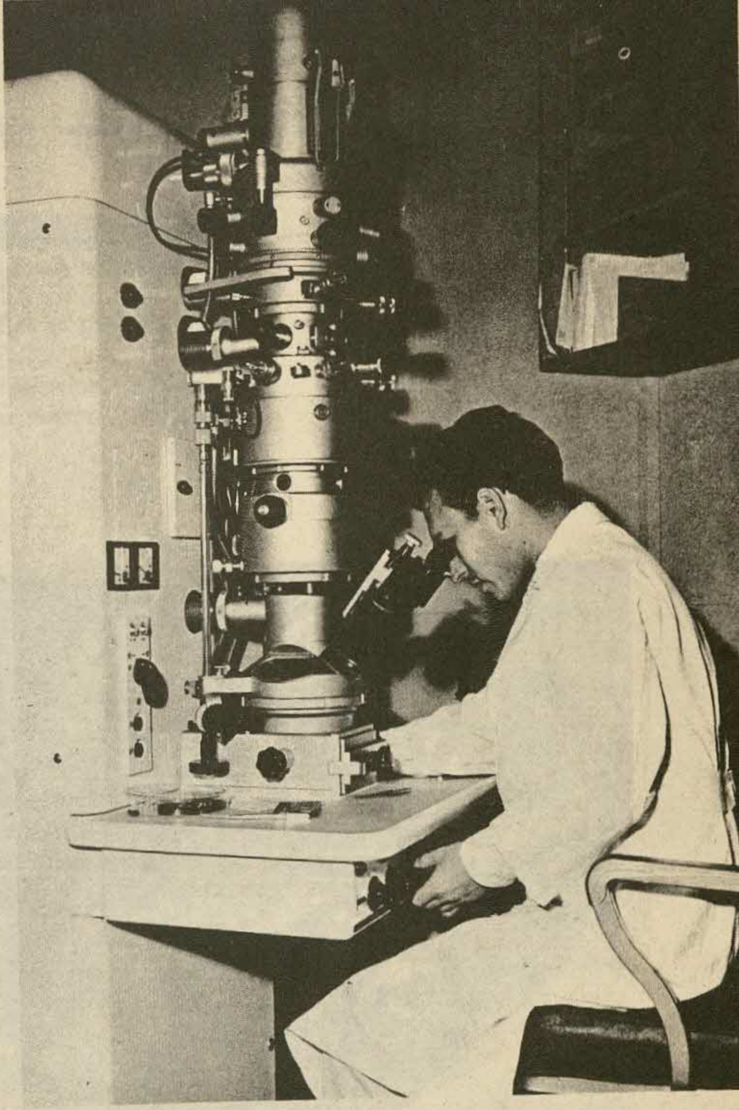




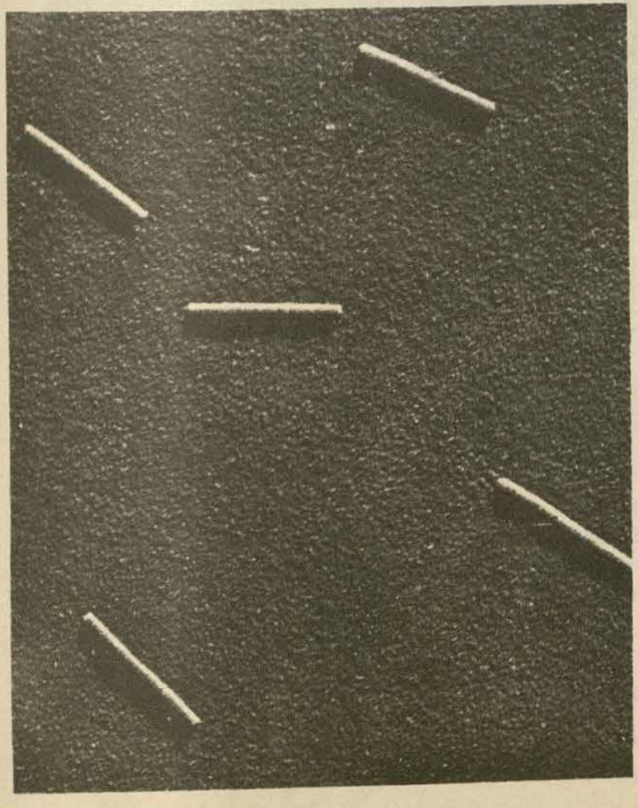
NGC 4594 Spiral nebula in Virgo, seen edge on; 200-in. photograph. (*Mount Wilson and Palomar Observatories*)

Observer in prime-focus cage changing film in the 200-in. Hale telescope. (*Mount Wilson and Palomar Observatories*)





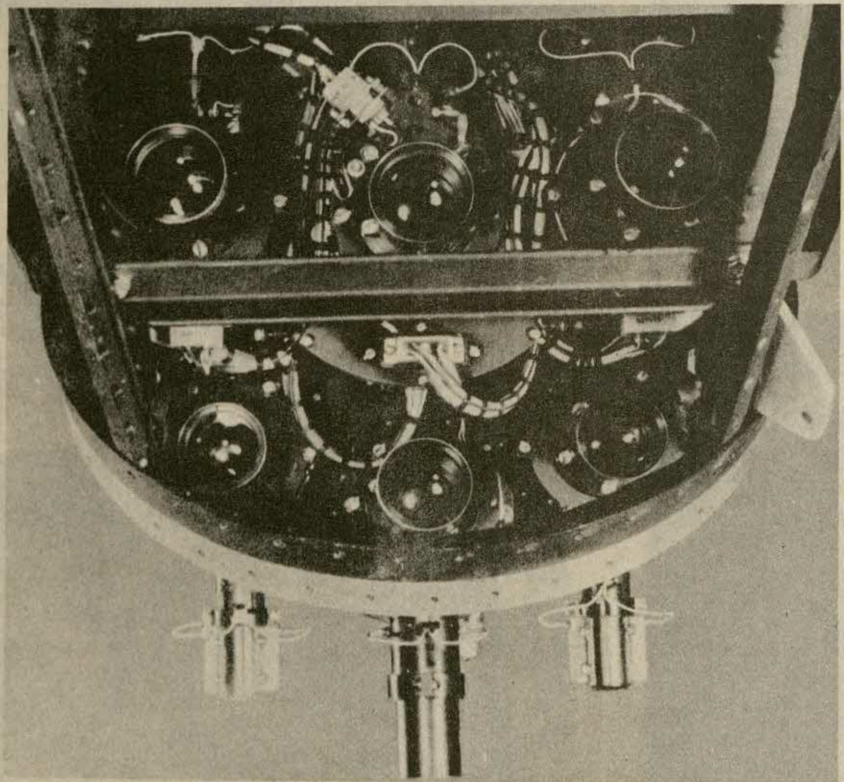
A Siemens electron microscope, operating at 50–100 kv and capable of a resolution of 10^{-7} cm. The stacked cylinders shown in the photograph contain magnetic lenses. The source of electrons is at the top and the final, magnified image is cast upon a fluorescent screen at the bottom. Photographic plates may be inserted in this plane to obtain a photographic record. Focusing is accomplished by varying the current in the magnetic lenses. (*Photograph by R. C. Williams, Virus Laboratory, University of California, Berkeley*)



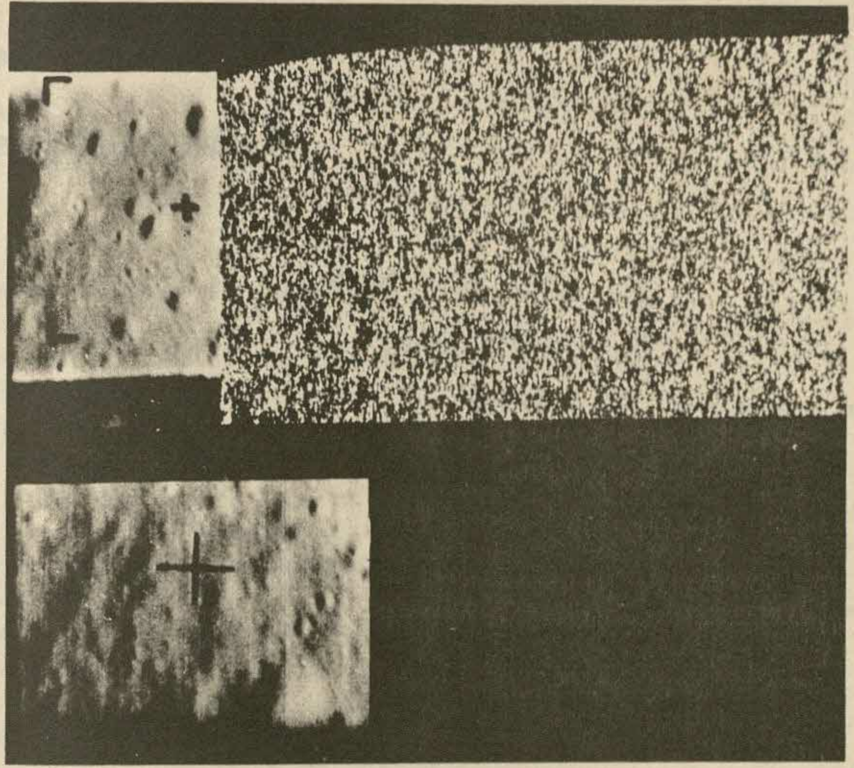
Date 17,2,2005
Accn. No 10978
Library

An electron photomicrograph of the particles of tobacco mosaic virus, at magnification 70,000. The rods are composed of protein and ribonucleic acid and are infectious when rubbed on leaves of tobacco plants. X-ray analysis shows that the particles are formed as a helix of about 2000 turns, with both the protein and the nucleic acid making up the gyres of the helix. (*Photograph by R. C. Williams, Virus Laboratory, University of California, Berkeley*)

Ranger Moon Shot. The Jet Propulsion Laboratory of the California Institute of Technology is responsible to the National Aeronautics and Space Administration for the Ranger program. On July 31, 1964, after a 68-hr flight, Ranger VII crashed on the moon. During the last 10 min of flight photographs were radioed back to earth as Ranger approached the lunar surface at 6000 mph. The black reseau marks on the photographs provide scale measurements.



Close-up of six television cameras (two for wide-angle views and four for narrow-angle views) carried by Ranger VII to photograph surface of the moon. (*NASA*)



Final frames: North at top. (*Above*) Altitude 1000 ft. This partial picture was the last taken by the P-3 camera with 25mm, f/1 lens before Ranger VII crashed onto the moon. The spacecraft was destroyed while transmitting, resulting in the receiver noise pattern at the right. The picture shows an area about 100 by 60 ft with a resolution 1000 times better than earth-based observations. The smallest craters are about 3 ft in diameter and 1 ft in depth.

(*Below*) Altitude about 3000 ft. The complete picture, showing an area about 100 ft on a side, was taken by the P-1 partial scan camera with 75mm, f/2 lens. Many of the craters shown have rounded shoulders in contrast with most larger lunar craters. (*NASA*)

S.C.E.R.T., W.B. LIBRARY
Date
Accn. No
Library

Chapter 2

Language and concepts 25
Vector notation 25
Equality of vectors 27
Vector addition 27
When is a physical quantity representable by a vector? 27
Vector derivatives 28
Example: Circular motion 30

Products of vectors 33
Scalar product of two vectors 33
Law of cosines 34
Direction cosines 35
Equation of a plane 35
Rate of doing work 36
Rate at which volume is swept out 36
Vector product 36
Area of a parallelogram 38
Volume of a parallelepiped 38
Law of sines 38
Triple products of vectors 39
Force on a particle in a magnetic field 39

Vectors in the Cartesian coordinate system 40
Invariance of A^2 *under rotation of the reference frame* 41
Examples: Various elementary vector operations 42

Useful vector identities 44

Further reading 44

Film list 45

Problems 45

Advanced Topic 1: Vectors and spherical polar coordinates 48

Advanced Topic 2: Crystal lattices and the reciprocal lattice 49

Mathematical Note 1: Equality of vectors in a spherical space 51

Mathematical Note 2: Generalized Cartesian vector notation 52

Historical Note: J. W. Gibbs 53

Vectors

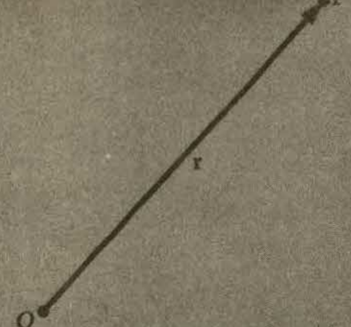
Language and Concepts. Language is an essential ingredient of abstract thought. It is difficult to think clearly and easily about sophisticated and abstract concepts in a language that has no words appropriate to such concepts. To express new scientific concepts new words are invented and added to languages; many such words are put together from classical Greek or Latin roots. If it satisfies the needs of the scientific community, a new word may be adopted in many modern languages. Thus *vector* in English is *vecteur* in French, *Vektor* in German, and вектор (pronounced "vector") in Russian. *A vector is a quantity having direction as well as magnitude.* This meaning for the word vector is a natural extension of an earlier usage, now obsolete, in astronomy: an imaginary straight line which joins a planet moving round a center or the focus of an ellipse, to that center or focus.

Vector Notation. Because symbols are the components of the language of mathematics, an important part of the art of mathematical analysis is the technique of using notation well. Vector notation has two great properties:

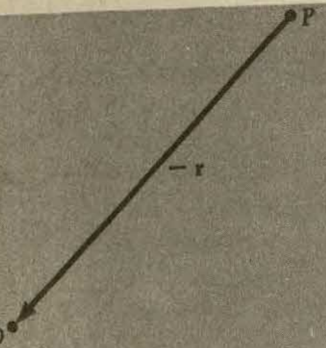
1. Formulation of a law of physics in terms of vectors is independent of the choice of coordinate axes. Vector notation offers a language in which statements have a physical content without ever introducing a coordinate system.
2. Vector notation is concise. Many physical laws have simple and transparent forms which are disguised when these laws are written in terms of a particular coordinate system.

Although in solving problems we may wish to work in special coordinate systems, we shall state the laws of physics in vector form wherever possible. Some of the more complicated laws, which cannot be expressed in vector form, may be expressed in terms of tensors. A *tensor* is a generalization of a vector and

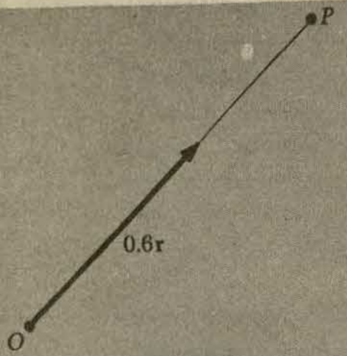
W.B. LIBRARY
Date
Accn. No
Library
Calcutta
25/3, B. C.



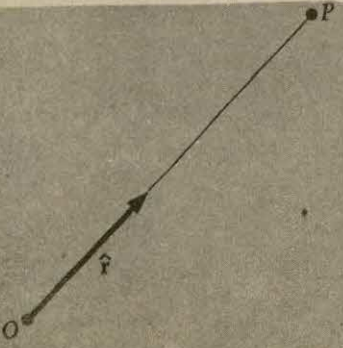
The vector **r** represents the position of a point *P* relative to another point *O* as origin.



The vector −**r** is equal in magnitude but opposite in direction to **r**.



The vector 0.6**r** is in the direction of **r** and is of magnitude 0.6*r*.



The vector $\hat{\mathbf{r}}$ is the unit vector in the direction of **r**. Note that $\mathbf{r} = \hat{\mathbf{r}}r$.

includes a vector as a special case. The vector analysis we know today is largely the result of work done toward the end of the nineteenth century by Josiah Willard Gibbs and Oliver Heaviside.

The vector notation we adopt is the following: On the blackboard a vector quantity named A is denoted by putting a wavy line under $\underset{\sim}{A}$ or by putting an arrow over the letter. The wavy line is a standard printer's mark—the printer will set $\underset{\sim}{A}$ as a boldface character **A**; in print vectors always appear in boldface type. The magnitude of a vector is printed in italics: *A* is the magnitude of **A**; *A* is also written as |**A**|. In a manuscript we underline a character if it is to be printed in italics. On the blackboard we usually do not underline an italic character when no ambiguity is introduced by the omission. A unit vector is a vector of unit length; a unit vector in the direction of **A** is written with a caret as $\hat{\mathbf{A}}$, which we read as "A hat," or "A caret." We summarize the notation by the identity

$$\mathbf{A} \equiv \hat{\mathbf{A}}A. \tag{1}$$

The usefulness and applicability of vectors to physical problems is largely based on Euclidean geometry. Statement of a law in terms of vectors usually carries with it the assumption of Euclidean geometry. If the geometry is not Euclidean, addition of two vectors in a simple and unambiguous way may not be possible. For curved space there exists a more general language, metric differential geometry, which is the language of general relativity, the domain of physics in which Euclidean geometry is no longer sufficiently valid.

We have considered a vector to be a quantity having direction as well as magnitude. This property has absolutely no reference to a particular coordinate system.† We shall see, however, that not all quantities having magnitude and direction are necessarily vectors. A *scalar* is defined to be any quantity whose value does not depend on a coordinate system. The magnitude of a vector is a scalar. The x coordinate of a fixed point is not a scalar, because the value of the x coordinate depends on the direction in which the x axis is chosen. Temperature T is a scalar; velocity **v** is a vector.

---

† We assume that the direction of a vector can be defined. For some purposes we might refer the direction to the laboratory room and for other purposes to the fixed stars.

Equality of Vectors. Two vectors **A** and **B** are defined to be equal if they have the same magnitude and direction. A vector does not necessarily have location, although a vector may refer to a quantity defined at a particular point. Two vectors can be compared even though they measure physical quantities defined at different points of space and time. If we did not have confidence, based on experiment, that we could consider space to be flat—except perhaps at enormous distances—then we could not unambiguously compare two vectors at different points (see Mathematical Note 1 at the end of Chap. 2).

Vector Addition. The sum of two vectors **A** and **B** is defined by the geometrical construction shown in the figures. This construction is often called the parallelogram law of addition of vectors. The sum $\mathbf{A} + \mathbf{B}$ is defined by carrying **B** parallel to itself until the tail of **B** coincides with the head of **A**. The vector drawn between the tail of **A** and the head of **B** is the sum $\mathbf{A} + \mathbf{B}$. From the figure it follows that $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$, so that vector addition is said to be commutative.

Vector subtraction is defined by the figures on page 28.

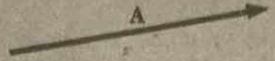
Vector addition satisfies the relation $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$, so that vector addition is said to be associative. The sum of a finite number of vectors is independent of the order in which they are added. If $\mathbf{A} - \mathbf{B} = \mathbf{C}$, then by adding **B** to both sides we obtain $\mathbf{A} = \mathbf{B} + \mathbf{C}$. In every way sums and differences of vectors can be manipulated as if they were numbers. If k is a scalar,

$$k(\mathbf{A} + \mathbf{B}) = k\mathbf{A} + k\mathbf{B}, \tag{2}$$

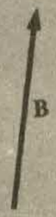
so that multiplication of a vector by a scalar is said to be distributive.

When is a physical quantity representable by a vector? We introduced the language of vectors to describe displacements in flat space. In addition to displacements there are other physical quantities which have the same laws of combination and the same invariance properties as displacements. Such quantities can also be represented by vectors. To be a vector a quantity must satisfy two conditions:

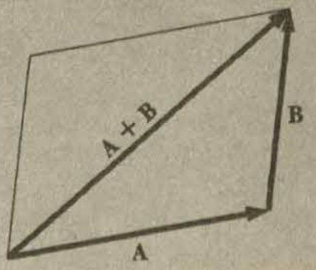
1. It must satisfy the parallelogram law of addition.
2. It must have a magnitude and a direction independent of the choice of coordinate system.



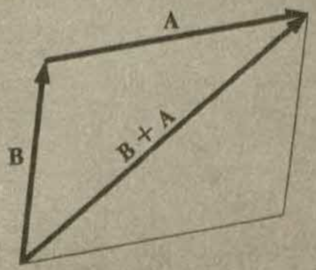
Vector **A**.



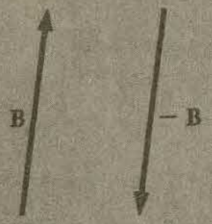
Vector **B**.



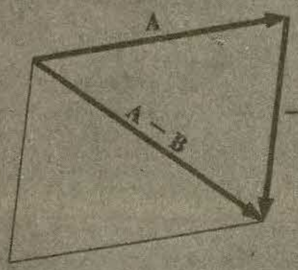
The vector sum $\mathbf{A} + \mathbf{B}$.



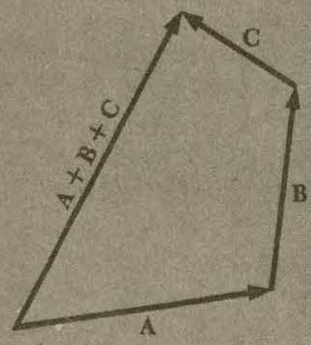
The vector sum $\mathbf{B} + \mathbf{A}$ is equal to $\mathbf{A} + \mathbf{B}$.



Vectors **B** and −**B**.



Formation of **A** − **B**; vector subtraction.



Sum of three vectors: **A** + **B** + **C**. Verify for yourself that this sum is equal to **B** + **A** + **C**.

Vector Derivatives. The velocity **v** of a particle is a vector; the acceleration **a** is also a vector. The *velocity* is the time rate of change of the position of a particle. The position of a particle at any time t can be specified by the vector $\mathbf{r}(t)$ which goes from a fixed point O to the particle. As the time advances, the particle moves and the position vector changes direction and magnitude. The difference between $\mathbf{r}(t_2)$ and $\mathbf{r}(t_1)$ is the difference between two vectors:

$$\Delta\mathbf{r} = \mathbf{r}(t_2) - \mathbf{r}(t_1) \tag{3}$$

and is itself a vector. If the vector **r** can be regarded as a function (a vector function) of the single scalar variable t, the value of $\Delta\mathbf{r}$ will be completely determined when the two values t_1 and t_2 are known. Thus in the accompanying figure $\Delta\mathbf{r}$ is the chord P_1P_2. The ratio

$$\frac{\Delta\mathbf{r}}{\Delta t}$$

is a vector collinear with the chord P_1P_2 but magnified in the ratio $1/\Delta t$. As Δt approaches zero, P_2 approaches P_1 and the chord P_1P_2 approaches the tangent at P_1. Then the vector

$$\frac{\Delta\mathbf{r}}{\Delta t} \text{ will approach } \frac{d\mathbf{r}}{dt},$$

which is a vector tangent to the curve at P_1 directed in the sense in which the variable t increases along the curve.

The vector

$$\frac{d\mathbf{r}}{dt} = \lim_{\Delta t \to 0} \frac{\Delta\mathbf{r}}{\Delta t} \tag{4}$$

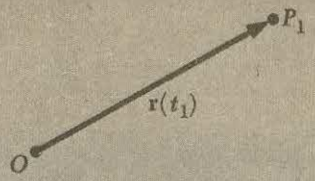
is called the *time derivative* of **r**. By definition the velocity is

$$\mathbf{v}(t) \equiv \frac{d\mathbf{r}}{dt}. \tag{5}$$

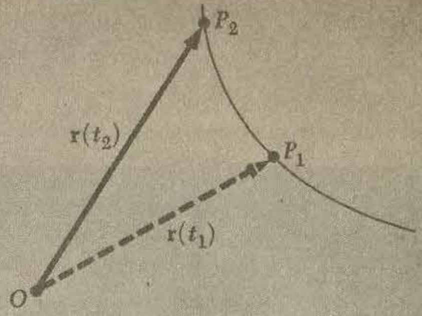
The magnitude $v = |\mathbf{v}|$ of the velocity is called the *speed* of the particle. The speed is a scalar.

Acceleration is also a vector; it is related to the velocity **v** just as **v** is related to **r**. We define the acceleration as

$$\mathbf{a} \equiv \frac{d\mathbf{v}}{dt} = \frac{d^2\mathbf{r}}{dt^2}. \tag{6}$$



The position P_1 of a particle at time t_1 is specified by the vector $\mathbf{r}(t_1)$ relative to the fixed origin at point O.



The particle has advanced to P_2 at time t_2.

Consider a particle moving in such a way that its position at any time t is given by the position vector $\mathbf{r}(t)$. We may write

$$\mathbf{r}(t) = r(t)\,\hat{\mathbf{r}}(t) \tag{7}$$

where the scalar $r(t)$ is the length of the vector and $\hat{\mathbf{r}}(t)$ is a vector of unit length in the direction of $\mathbf{r}$. The derivative of $\mathbf{r}(t)$ is defined as

$$\frac{d\mathbf{r}}{dt} = \frac{d}{dt}r(t)\,\hat{\mathbf{r}}(t) = \lim_{\Delta t\to 0}\frac{r(t+\Delta t)\,\hat{\mathbf{r}}(t+\Delta t) - r(t)\,\hat{\mathbf{r}}(t)}{\Delta t}. \tag{8}$$

We may rewrite the numerator to obtain

$$\left[r(t) + \frac{dr}{dt}\Delta t\right]\left[\hat{\mathbf{r}}(t) + \frac{d\hat{\mathbf{r}}}{dt}\Delta t\right] - r(t)\,\hat{\mathbf{r}}(t)$$
$$= \Delta t\left[\frac{dr}{dt}\hat{\mathbf{r}} + r\frac{d\hat{\mathbf{r}}}{dt}\right] + (\Delta t)^2\left[\frac{dr}{dt}\frac{d\hat{\mathbf{r}}}{dt}\right]. \tag{9}$$

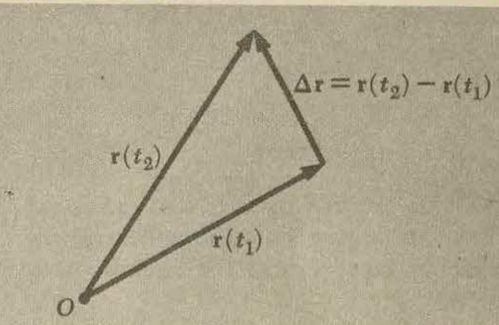
In the limit as $\Delta t \to 0$ the last term on the right-hand side may be neglected and we have

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = \frac{dr}{dt}\hat{\mathbf{r}} + r\frac{d\hat{\mathbf{r}}}{dt}. \tag{10}$$

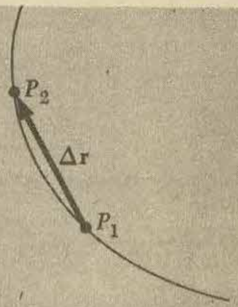
This is an example of the general rule for differentiation of the product of a scalar $a(t)$ and a vector $\mathbf{b}(t)$

$$\frac{d}{dt}a\mathbf{b} = \frac{da}{dt}\mathbf{b} + a\frac{d\mathbf{b}}{dt}. \tag{11}$$

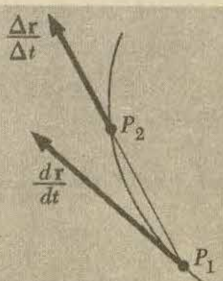
One contribution to the velocity in (10) comes from the change in the direction $\hat{\mathbf{r}}$; the other contribution comes from the change in the length r.



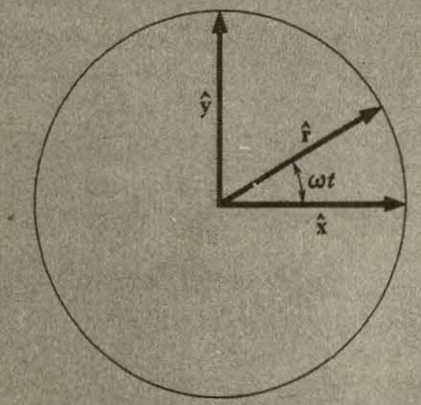
The vector $\Delta\mathbf{r}$ is the difference between $\mathbf{r}(t_2)$ and $\mathbf{r}(t_1)$.



$\Delta\mathbf{r}$ is the chord between the points P_1 and P_2 on the trajectory of the particle.



As $\Delta t = t_2 - t_1 \to 0$, the vector $\Delta\mathbf{r}/\Delta t$ collinear with the chord approaches the velocity vector $d\mathbf{r}/dt$ collinear with the tangent to the trajectory at point P_1.



Particle moving in a circle of unit radius and with angular frequency ω. The particle velocity is given by Eq. (19) and the acceleration by Eq. (22).

Example. Circular motion. This example is extremely important! We want to obtain explicit expressions for the velocity and acceleration of a particle moving at constant speed in a circular orbit of constant radius r. A circular orbit can be described by

$$\mathbf{r}(t) = r\hat{\mathbf{r}}(t), \tag{12}$$

provided that r is constant and that the unit vector $\hat{\mathbf{r}}$ rotates at a constant rate. We can represent such a unit vector by

$$\hat{\mathbf{r}}(t) = \cos\omega t\,\hat{\mathbf{x}} + \sin\omega t\,\hat{\mathbf{y}}, \tag{13}$$

where $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are perpendicular constant unit vectors. Here ω, the Greek character omega, is a constant and is called the *angular frequency* or *angular velocity* of the motion. It is measured in radians per unit time. The vector $\hat{\mathbf{r}}$ rotates counterclockwise if ω is positive, and after a time t the vector $\hat{\mathbf{r}}$ makes the angle ωt radians with the x direction. (Remember that there are 2π radians in 360°). These statements about $\hat{\mathbf{r}}$ follow directly from the trigonometrical definitions of the cosine and sine functions. Observe that at $t = 0$ the unit vector $\hat{\mathbf{r}}$ is along the x axis.

As illustration, consider a value of the time such that $\omega t = \frac{1}{4}\pi$ radians, which is 45°. We know that $\cos\frac{1}{4}\pi = \sin\frac{1}{4}\pi = 1/\sqrt{2}$, so that

$$\hat{\mathbf{r}} = \frac{1}{\sqrt{2}}\hat{\mathbf{x}} + \frac{1}{\sqrt{2}}\hat{\mathbf{y}}, \tag{14}$$

which is a unit vector at an angle of 45° counterclockwise from the x axis. At the later time such that $\omega t = \frac{1}{2}\pi$ radians, which is 90°, we have $\cos\frac{1}{2}\pi = 0$; $\sin\frac{1}{2}\pi = 1$, so that

$$\hat{\mathbf{r}} = \hat{\mathbf{y}}. \tag{15}$$

The unit vector is now along the y direction.

To obtain the velocity of the particle in circular motion we employ the result (10), but with $dr/dt = 0$ because the radius r is constant for a circle. Thus, from (12) and (13),

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = r\frac{d\hat{\mathbf{r}}}{dt} = r\left(\hat{\mathbf{x}}\frac{d}{dt}\cos\omega t + \hat{\mathbf{y}}\frac{d}{dt}\sin\omega t\right). \tag{16}$$

We need the derivatives of the cosine and sine. We recall from the calculus that

$$\frac{d}{dt}\sin t = \cos t; \qquad \frac{d}{dt}\sin \omega t = \omega \cos \omega t; \tag{17}$$

$$\frac{d}{dt}\cos t = -\sin t; \qquad \frac{d}{dt}\cos \omega t = -\omega \sin \omega t. \tag{18}$$

It is not at all difficult to derive these results. If you have not seen them before, study the note.

With the help of (17) and (18) we may rewrite (16) as

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = r(-\omega \sin \omega t\,\hat{\mathbf{x}} + \omega \cos \omega t\,\hat{\mathbf{y}}). \tag{19}$$

The magnitude of the velocity is ωr, as may be seen by forming (using the definition of scalar product given in (28) below, and the property $\hat{\mathbf{x}} \cdot \hat{\mathbf{y}} = 0$)

$$\begin{aligned} v^2 = \mathbf{v} \cdot \mathbf{v} &= \omega^2 r^2(-\sin \omega t\,\hat{\mathbf{x}} + \cos \omega t\,\hat{\mathbf{y}}) \cdot (-\sin \omega t\,\hat{\mathbf{x}} + \cos \omega t\,\hat{\mathbf{y}}) \\ &= \omega^2 r^2(\sin^2 \omega t + \cos^2 \omega t) = \omega^2 r^2. \end{aligned} \tag{20}$$

We have used the identity $\sin^2 + \cos^2 \equiv 1$. This establishes the important result that the speed of a particle in uniform circular motion is

$$\boxed{v = \omega r.} \tag{21}$$

The acceleration associated with circular motion is found on differentiation of (19):

$$\mathbf{a} = \frac{d\mathbf{v}}{dt} = r(-\omega^2 \cos \omega t\,\hat{\mathbf{x}} - \omega^2 \sin \omega t\,\hat{\mathbf{y}}). \tag{22}$$

We recognize by comparison with (12) and (13) that the right-hand side of (22) is just $-\omega^2\mathbf{r}$:

$$\mathbf{a} = -\omega^2 r(\cos \omega t\,\hat{\mathbf{x}} + \sin \omega t\,\hat{\mathbf{y}}) = -\omega^2 \mathbf{r}. \tag{23}$$

Thus the magnitude of the acceleration in circular motion is

$$\boxed{a = \omega^2 r,} \tag{24}$$

and the direction of the acceleration is in the direction $-\mathbf{r}$, which is toward the center of the circle. Using $v = \omega r$ from (21), we may rewrite (24) as

$$\boxed{a = \frac{v^2}{r}.} \tag{25}$$

This acceleration is called the *centripetal* (center-seeking) accel-

Note: The derivative of sin t with respect to t is defined in the usual way as

$$\begin{aligned} \frac{d}{dt}\sin t &= \lim_{\Delta t \to 0} \frac{\sin (t + \Delta t) - \sin t}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{\sin t \cos \Delta t + \cos t \sin \Delta t - \sin t}{\Delta t}, \end{aligned}$$

where we have used the standard trigonometric identity

$$\sin (a + b) = \sin a \cos b + \cos a \sin b.$$

Now from the geometrical definitions of cosine and sine it is obvious that

$$\lim_{\Delta t \to 0} \cos \Delta t = 1; \qquad \lim_{\Delta t \to 0} \frac{\sin \Delta t}{\Delta t} = 1.$$

On substituting these results above we have

$$\frac{d}{dt}\sin t = \cos t.$$

The reader may proceed similarly to obtain

$$\frac{d}{dt}\cos t.$$

eration, and it will be familiar to you from high school physics.

The angular frequency ω has a simple connection with the ordinary frequency f. In unit time the vector $\hat{\mathbf{r}}$ in (13) sweeps out ω radians, so that ω denotes the number of radians swept out per unit time. But the ordinary frequency f is defined as the number of cycles of motion swept out per unit time. There being 2π radians in one cycle, we must have

$$\boxed{2\pi f = \omega.} \tag{26}$$

The *period* T of the motion is defined as the time to complete one cycle. We see from (13) that one cycle is completed at a time T such that $\omega T = 2\pi$, or

$$\boxed{T = \frac{2\pi}{\omega} = \frac{1}{f}.} \tag{27}$$

For your numerical orientation, suppose that the frequency f is 60 cycles per second, or 60 cps. Then the period

$$T = \frac{1}{f} = \frac{1}{60} \cong 0.017 \text{ sec},$$

and the angular frequency is

$$\omega = 2\pi f \cong 377 \text{ radians/sec}.$$

If the radius of a circular orbit is 10 cm, then the velocity is

$$v = \omega r \cong (377)(10) \simeq 3.8 \times 10^3 \text{ cm/sec}.$$

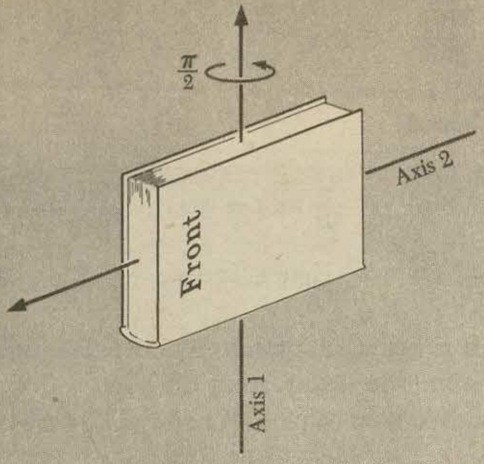
The acceleration at any point of the orbit is

$$a = \omega^2 r \simeq (3.8 \times 10^2)^2(10) \simeq 1 \times 10^6 \text{ cm/sec}^2.$$

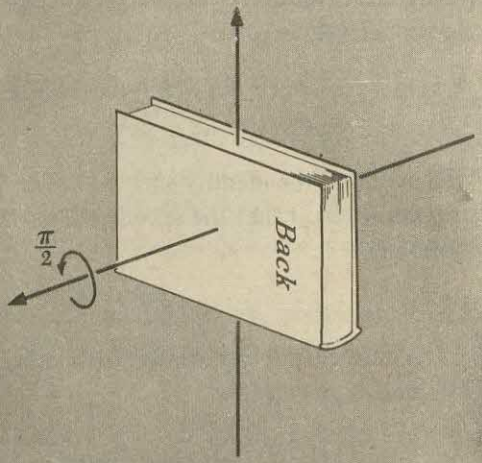
In Chap. 3 a numerical example is worked out which shows that the acceleration of a point fixed on the surface of the earth at the equator, due to the rotation of the earth about its own axis, is about 3.4 cm/sec^2.

The definitions of **r**, **v**, and **a** ensure that they are vectors. The force **F**, the electric field **E**, and the magnetic field **B** are also vectors, but we must look to experiment to verify their required vector properties.

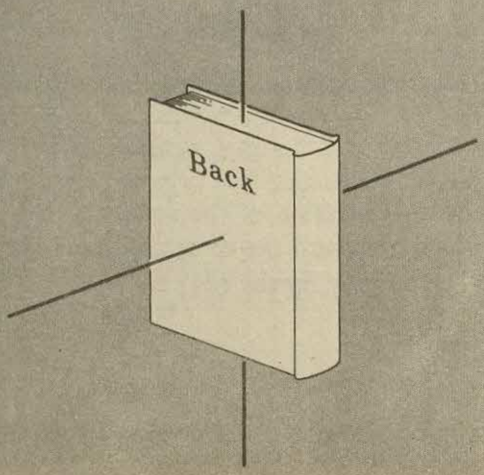
By experiment $\mathbf{F} = M\mathbf{a}$, where the mass M is a constant scalar.†



Original orientation of book. It is then rotated by $\pi/2$ radians about Axis 1.



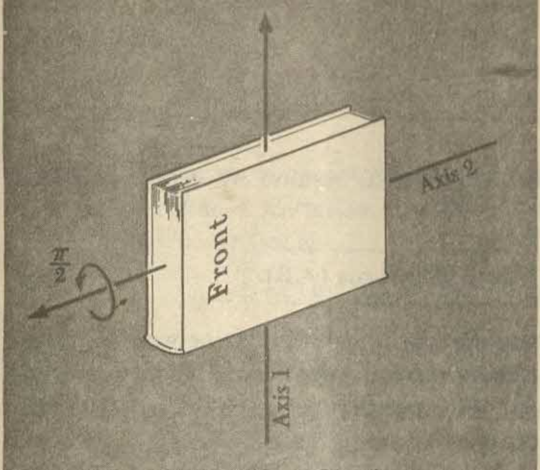
Orientation after a rotation of $\pi/2$ radians about Axis 1.



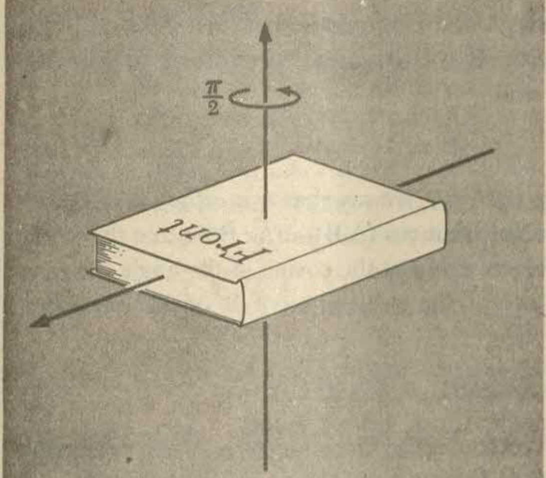
Orientation after a subsequent rotation of $\pi/2$ radians about Axis 2.

---

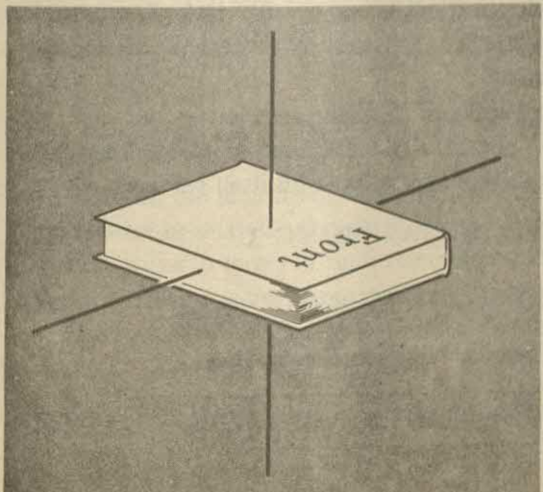
† If M is not constant, $\mathbf{F} = \frac{d}{dt}(M\mathbf{v}) = \frac{dM}{dt}\mathbf{v} + M\frac{d\mathbf{v}}{dt} = \frac{dM}{dt}\mathbf{v} + M\mathbf{a}$.



Original orientation of book.



Orientation after a rotation of $\pi/2$ radians about Axis 2.



Orientation after subsequent rotation of $\pi/2$ radians about Axis 1.

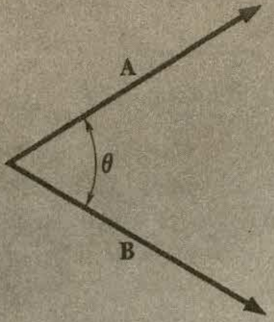
Because **a** is a vector, **F** must also be a vector. The electric field intensity **E** is defined as the force in an electric field on a particle of unit charge at rest; thus **E** must be a vector. It is known experimentally that magnetic fields add vectorially: The net effect of $\mathbf{B}_1$ and $\mathbf{B}_2$ acting together is exactly equivalent to that of a single magnetic field $\mathbf{B}_1 + \mathbf{B}_2$, and **B** is a vector.

Not all quantities that have magnitude and direction are necessarily vectors. For example, a rotation of a rigid body about a particular axis fixed in space has a magnitude (the angle of rotation) and a direction (the direction of the axis). But two such rotations do not combine according to the vector law of addition, unless the angles of rotation are infinitesimally small. This is easily seen when the two axes are perpendicular to each other and the rotations are by $\pi/2$ radians (90 deg). Consider the object (a book). The rotation (1) leaves it as in the second figure, and a subsequent rotation (2) about another axis leaves the object as in the third figure. But if to the object as originally oriented we apply first the rotation (2) and then the rotation (1), the object ends up as shown in the last figure. The orientation in the sixth figure is *not* the same as in the third. Obviously the commutative law of addition is not satisfied by these rotations. Despite the fact that they have a magnitude and a direction, finite rotations cannot be represented as vectors.

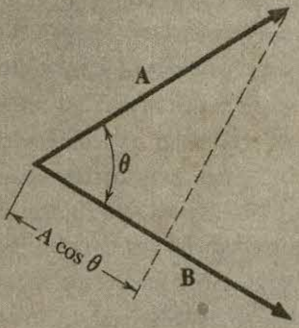
Products of Vectors

There are two particularly useful ways in which to define the product of two vectors. Both products satisfy the distributive law of multiplication: the product of **A** into the sum of **B** + **C** is equal to the sum of the products of **A** into **B** and **A** into **C**. One type of product is a scalar, the other is for many purposes a vector. Both products are useful in physics. Other possible definitions of product are not useful: why is AB not a useful definition of the product of two vectors? By AB we mean the ordinary product, $|\mathbf{A}|\,|\mathbf{B}|$, of the magnitudes of **A** and **B**. We observe that if $\mathbf{D} = \mathbf{B} + \mathbf{C}$, then in general $AD \neq AB + AC$. This absence of the distributive property makes AB useless as a product of **A** and **B**.

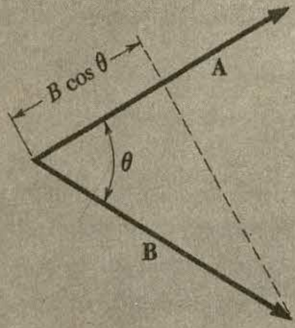
Scalar Product of Two Vectors. The scalar product of **A** and **B** is defined as that number which is obtained by taking the magnitude of **A** times the magnitude of **B** times the cosine of the angle between them. The scalar product is a scalar. Often the scalar



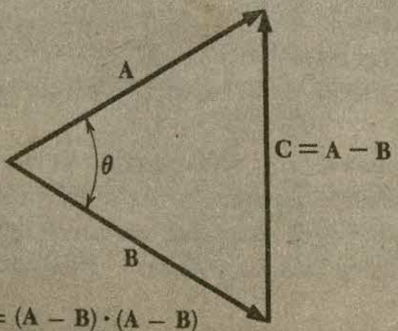
In forming $\mathbf{A}\cdot\mathbf{B}$, bring vectors **A** and **B** to a common origin.



$B(A\cos\theta) = \mathbf{A}\cdot\mathbf{B}$.



$A(B\cos\theta) = \mathbf{A}\cdot\mathbf{B}$. Here θ, the Greek character theta, denotes the angle between **A** and **B**.



$$\begin{aligned}\mathbf{C}\cdot\mathbf{C} = C^2 &= (\mathbf{A}-\mathbf{B})\cdot(\mathbf{A}-\mathbf{B})\\ &= A^2 + B^2 - 2\mathbf{A}\cdot\mathbf{B}\\ &= A^2 + B^2 - 2AB\cos\theta.\end{aligned}$$

product is called the dot product, because we denote it by the symbol

$$\boxed{\mathbf{A}\cdot\mathbf{B} \equiv AB\cos(\mathbf{A},\mathbf{B}).} \tag{28}$$

Here $\cos(\mathbf{A},\mathbf{B})$ denotes the cosine of the angle between **A** and **B**. We see that no coordinate system is involved at all in the definition of scalar product. We note that $\cos(\mathbf{A},\mathbf{B}) = \cos(\mathbf{B},\mathbf{A})$, so that the scalar product is commutative:

$$\mathbf{A}\cdot\mathbf{B} = \mathbf{B}\cdot\mathbf{A}. \tag{29}$$

We read $\mathbf{A}\cdot\mathbf{B}$ as "**A** *dot* **B**."

If the angle between **A** and **B** should lie between $\pi/2$ and $3\pi/2$, then $\cos(\mathbf{A},\mathbf{B})$ and $\mathbf{A}\cdot\mathbf{B}$ will be negative numbers. If $\mathbf{A} = \mathbf{B}$, then $\cos(\mathbf{A},\mathbf{B}) = 1$ and

$$\mathbf{A}\cdot\mathbf{B} = A^2 = |\mathbf{A}|^2. \tag{30}$$

If $\mathbf{A}\cdot\mathbf{B} = 0$ and $A \neq 0$, $B \neq 0$, we say that **A** is *orthogonal* to **B** or perpendicular to **B**. Note that $\cos(\mathbf{A},\mathbf{B}) = \hat{\mathbf{A}}\cdot\hat{\mathbf{B}}$, so that the scalar product of two unit vectors is just the cosine of the angle between them. The magnitude of the projection of **B** on the direction of **A** is

$$B\cos(\mathbf{A},\mathbf{B}) = B\hat{\mathbf{A}}\cdot\hat{\mathbf{B}} = \mathbf{B}\cdot\hat{\mathbf{A}}, \tag{31}$$

where $\hat{\mathbf{A}}$ is the unit vector in the direction of **A**. The projection of **A** on the direction of **B** is

$$A\cos(\mathbf{A},\mathbf{B}) = \mathbf{A}\cdot\hat{\mathbf{B}}. \tag{32}$$

Scalar product multiplication has no inverse: if $\mathbf{A}\cdot\mathbf{X} = b$, there is no unique solution for **X**. Division by a vector is a meaningless, undefined operation.

We treat several applications of scalar products:

1. *Law of cosines.* Let $\mathbf{A} - \mathbf{B} = \mathbf{C}$; then on taking the scalar product of each side of this expression with itself we have

$$(\mathbf{A}-\mathbf{B})\cdot(\mathbf{A}-\mathbf{B}) = \mathbf{C}\cdot\mathbf{C} \tag{33}$$

or

$$A^2 + B^2 - 2\mathbf{A}\cdot\mathbf{B} = C^2 \tag{34}$$

which is exactly the famous trigonometric relation

$$\boxed{A^2 + B^2 - 2AB\cos(\mathbf{A},\mathbf{B}) = C^2.} \tag{35}$$

2. *Direction cosines.* Let $\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}$ be three orthogonal† unit vectors which define a Cartesian coordinate system. An arbitrary vector **A** may be written as

$$\mathbf{A} = \hat{\mathbf{x}}\,(\mathbf{A}\cdot\hat{\mathbf{x}}) + \hat{\mathbf{y}}\,(\mathbf{A}\cdot\hat{\mathbf{y}}) + \hat{\mathbf{z}}\,(\mathbf{A}\cdot\hat{\mathbf{z}}). \tag{36}$$

Here $(\mathbf{A}\cdot\hat{\mathbf{x}})$, $(\mathbf{A}\cdot\hat{\mathbf{y}})$, and $(\mathbf{A}\cdot\hat{\mathbf{z}})$ are called the components of A and are often denoted by A_x, A_y, and A_z, respectively. This relation may be verified by forming the scalar product of both sides with $\hat{\mathbf{x}}$ and using (by the definition of orthogonal unit vectors) $\hat{\mathbf{x}}\cdot\hat{\mathbf{x}} = 1$; $\hat{\mathbf{x}}\cdot\hat{\mathbf{y}} = 0$; $\hat{\mathbf{x}}\cdot\hat{\mathbf{z}} = 0$, so that

$$\begin{aligned}(\mathbf{A}\cdot\hat{\mathbf{x}}) &= (\hat{\mathbf{x}}\cdot\hat{\mathbf{x}})\,(\mathbf{A}\cdot\hat{\mathbf{x}}) + (\hat{\mathbf{y}}\cdot\hat{\mathbf{x}})(\mathbf{A}\cdot\hat{\mathbf{y}}) + (\hat{\mathbf{z}}\cdot\hat{\mathbf{x}})(\mathbf{A}\cdot\hat{\mathbf{z}})\\ &= (\mathbf{A}\cdot\hat{\mathbf{x}}).\end{aligned} \tag{37}$$

We may write $\hat{\mathbf{A}}$ as the unit vector in the direction of **A** in (36) as

$$\hat{\mathbf{A}} = \hat{\mathbf{x}}\cos(\hat{\mathbf{A}},\hat{\mathbf{x}}) + \hat{\mathbf{y}}\cos(\hat{\mathbf{A}},\hat{\mathbf{y}}) + \hat{\mathbf{z}}\cos(\hat{\mathbf{A}},\hat{\mathbf{z}}). \tag{38}$$

The three cosines in this equation are called the *direction cosines* of $\hat{\mathbf{A}}$ or the direction cosines of **A**, referred to the unit basis vectors $\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}$ of the orthogonal Cartesian coordinate system. On forming the scalar product of each side of (38) with itself we find another well-known relation:

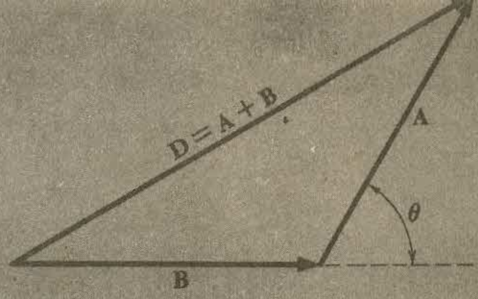
$$\boxed{1 = \cos^2(\hat{\mathbf{A}},\hat{\mathbf{x}}) + \cos^2(\hat{\mathbf{A}},\hat{\mathbf{y}}) + \cos^2(\hat{\mathbf{A}},\hat{\mathbf{z}}).} \tag{39}$$

The sum of the squares of the three direction cosines is equal to unity.

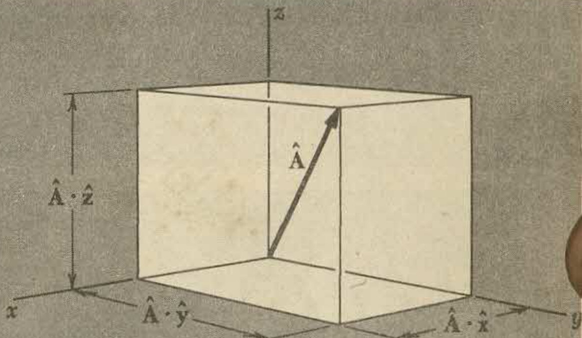
3. *Equation of a plane.* Let **N** be a normal to the plane under consideration, which is drawn from an origin O not in the plane. Let **r** be an arbitrary vector from the origin O to any point P in the plane. The projection of **r** on **N** must be equal in magnitude to N. Thus the plane is described by the equation

$$\boxed{\mathbf{r}\cdot\mathbf{N} = N^2.} \tag{40}$$

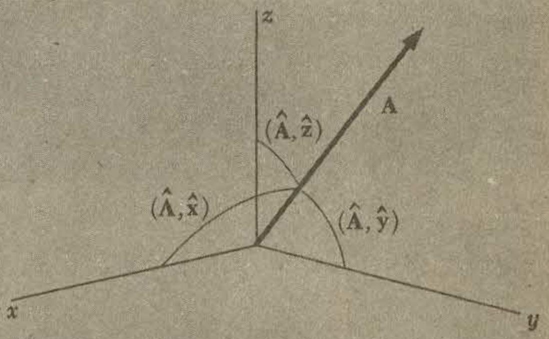
To establish the identity of this compact expression with the usual expression in analytic geometry for the equation of a plane we write **N** and **r** in terms of their components N_x, N_y, N_z, and x, y, z in the Cartesian coordinate system defined by the orthogonal unit vectors $\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}$. Now Eq. (40) assumes the form



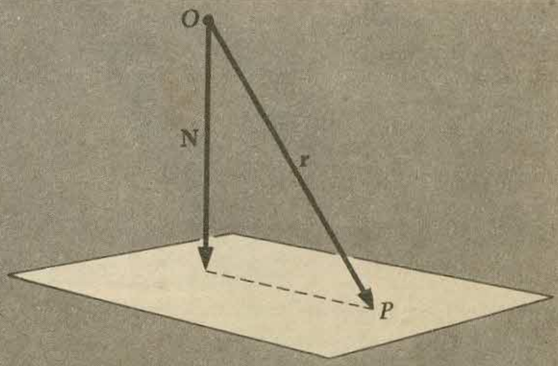
$$\begin{aligned}\mathbf{D}\cdot\mathbf{D} = D^2 &= (\mathbf{A}+\mathbf{B})\cdot(\mathbf{A}+\mathbf{B})\\ &= A^2 + B^2 + 2AB\cos\theta\end{aligned}$$



Direction cosines $\hat{\mathbf{A}}\cdot\hat{\mathbf{z}}$; $\hat{\mathbf{A}}\cdot\hat{\mathbf{y}}$; $\hat{\mathbf{A}}\cdot\hat{\mathbf{x}}$ shown as projections.



Direction cosines refer to the angles indicated.



Equation of a plane; **N** is the normal to the plane from the origin O. The equation of the plane is $\mathbf{N}\cdot\mathbf{r} = N^2$.

---

† *Orthogonal* as used here means mutually perpendicular.

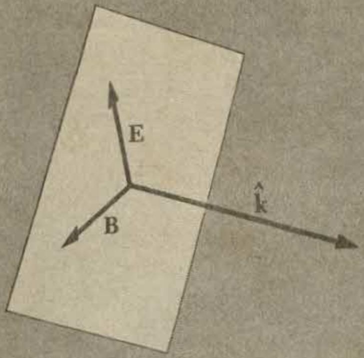
$$(\hat{\mathbf{x}}x + \hat{\mathbf{y}}y + \hat{\mathbf{z}}z) \cdot (\hat{\mathbf{x}}N_x + \hat{\mathbf{y}}N_y + \hat{\mathbf{z}}N_z) = N^2, \tag{41}$$

or

$$xN_x + yN_y + zN_z = N^2. \tag{42}$$

If $\hat{\mathbf{k}}$ is the unit vector in the direction of propagation of a plane electromagnetic wave in free space, then (as we shall see in Vol. III) the electric and magnetic intensity vectors **E** and **B** must lie in a plane normal to $\hat{\mathbf{k}}$. We can express the geometrical condition by the relations

$$\hat{\mathbf{k}} \cdot \mathbf{E} = 0; \qquad \hat{\mathbf{k}} \cdot \mathbf{B} = 0. \tag{43}$$



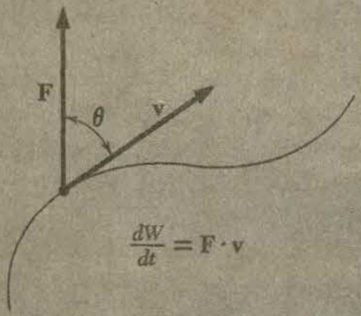
Electric and magnetic fields in a plane electromagnetic wave in free space are perpendicular to the propagation direction $\hat{\mathbf{k}}$. Thus $\hat{\mathbf{k}} \cdot \mathbf{E} = \hat{\mathbf{k}} \cdot \mathbf{B} = 0$.

4. Rate of doing work. In elementary physics (also see Chap. 5) we learned that the rate at which a force **F** does work on a particle moving with velocity **v** is equal to $Fv \cos(\mathbf{F},\mathbf{v})$. We recognize this expression as just the scalar product

$$\mathbf{F} \cdot \mathbf{v}.$$

If we write generally the derivative dW/dt as a symbol for the rate of doing work, then

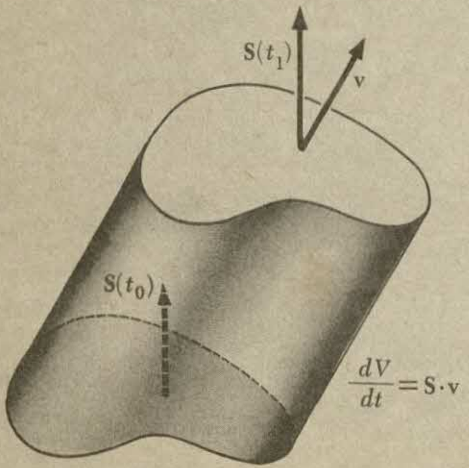
$$\boxed{\frac{dW}{dt} = \mathbf{F} \cdot \mathbf{v}.} \tag{44}$$



Rate at which a force **F** does work on a particle moving with velocity **v**.

5. Rate at which volume is swept out. Let **S** be a vector normal to a plane area of magnitude S, and let **v** denote the velocity at which the area is moved. We recognize that the volume per unit time traversed by the area **S** is a cylinder of base area S and slant height v. The rate dV/dt at which volume is swept out is

$$\frac{dV}{dt} = \mathbf{S} \cdot \mathbf{v}. \tag{45}$$

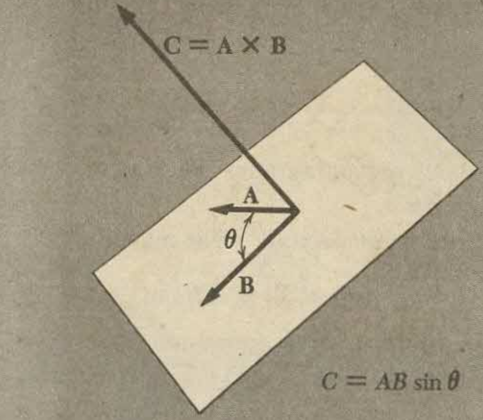


Rate dV/dt at which area **S** moving with velocity **v** sweeps out volume.

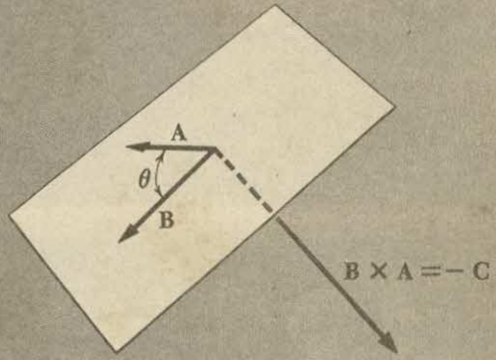
Vector Product. Another type of product of two vectors is used widely in physics. This product is a vector rather than a scalar, but it is a vector in a somewhat restricted sense. The *vector product* $\mathbf{A} \times \mathbf{B}$ is defined to be a vector normal to the plane that includes **A** and **B** and which has a magnitude $AB\,|\sin(\mathbf{A},\mathbf{B})|$:

$$\boxed{\mathbf{C} = \mathbf{A} \times \mathbf{B} = \hat{\mathbf{C}}\,AB\,|\sin(\mathbf{A},\mathbf{B})|\ .} \tag{46}$$

We read $\mathbf{A} \times \mathbf{B}$ as "**A** *cross* **B**." The sense of **C** is determined as a matter of fixed convention by the *right-hand-thread* rule: The vector **A** in the first position in the product is rotated by the



Vector product $\mathbf{C} = \mathbf{A} \times \mathbf{B}$.



Vector product $\mathbf{B} \times \mathbf{A}$ is opposite to $\mathbf{A} \times \mathbf{B}$.

smallest angle that will bring it into coincidence with the direction of **B**. The sense of **C** is that of the direction of motion of a screw with a right-hand thread (the standard thread in the United States), when the screw is turned in the same direction as was the vector **A**.

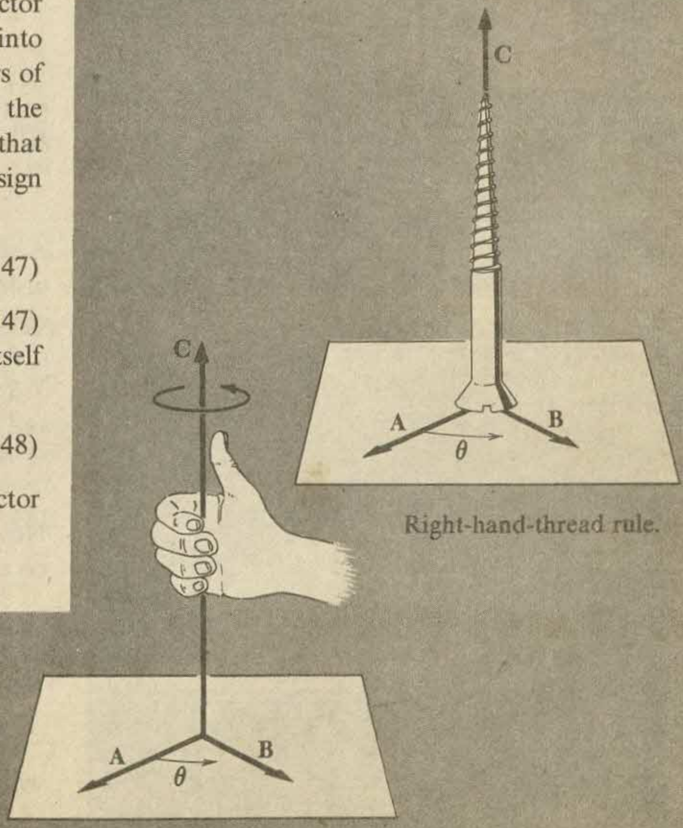
Let us state the rule for the direction of **C** in another way: First, place together the tails of vectors **A** and **B**—this defines a plane. Vector **C** is perpendicular to this plane; that is, the vector product $\mathbf{A} \times \mathbf{B}$ is perpendicular to both **A** and **B**. Rotate **A** into **B** through the lesser of the two possible angles—curl the fingers of the right hand in the direction in which **A** was rotated, and the thumb will point in the direction of $\mathbf{C} = \mathbf{A} \times \mathbf{B}$. Note that because of this sign convention $\mathbf{B} \times \mathbf{A}$ is a vector opposite in sign to $\mathbf{A} \times \mathbf{B}$:

$$\mathbf{B} \times \mathbf{A} = -\mathbf{A} \times \mathbf{B}. \tag{47}$$

Thus the vector product is not commutative. It follows from (47) that $\mathbf{A} \times \mathbf{A} = 0$, so that the vector product of a vector with itself is zero. The vector product does obey the distributive law:

$$\mathbf{A} \times (\mathbf{B} + \mathbf{C}) = \mathbf{A} \times \mathbf{B} + \mathbf{A} \times \mathbf{C}. \tag{48}$$

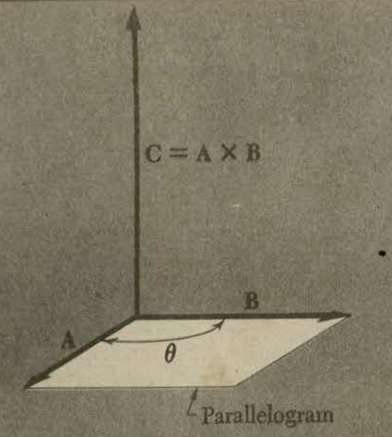
The somewhat tedious proof may be found in any book on vector analysis.



Right-hand-thread rule.

Same rule, applied with right hand.

In the following paragraphs we treat several applications of vector products.

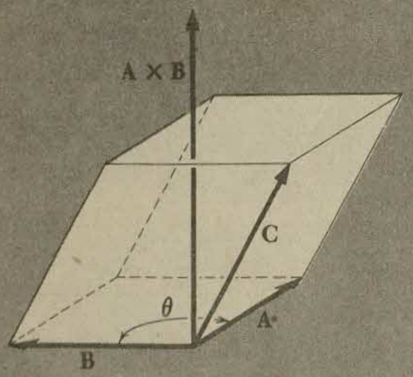


Vector area of parallelogram is $\mathbf{C} = \mathbf{A} \times \mathbf{B} = AB\,|\sin\theta|\,\hat{\mathbf{C}}$.

1. Area of a parallelogram. The magnitude

$$|\mathbf{A} \times \mathbf{B}| = AB\,|\sin(\mathbf{A},\mathbf{B})| \tag{49}$$

is the area of the parallelogram with sides **A** and **B** (or twice the area of the triangle with sides **A** and **B**). The direction of $\mathbf{A} \times \mathbf{B}$ is normal to the plane of the parallelogram; thus we may think of $\mathbf{A} \times \mathbf{B}$ as the *vector area* of the parallelogram. Because we have given signs to the sides **A** and **B**, the vector area comes endowed with a direction. There are physical applications where it is convenient to be able to give a direction to an area.



$\mathbf{A} \times \mathbf{B} \cdot \mathbf{C}$ = base area × height = volume of parallelepiped.

2. Volume of a parallelepiped. The scalar

$$\boxed{|(\mathbf{A} \times \mathbf{B}) \cdot \mathbf{C}| = V} \tag{50}$$

is the volume of the parallelepiped of which $\mathbf{A} \times \mathbf{B}$ is the area of the base and **C** the slant height or edge. If the three vectors **A**, **B**, **C** lie in the same plane, the volume will be zero; thus three vectors are coplanar if, and only if, $(\mathbf{A} \times \mathbf{B}) \cdot \mathbf{C} = 0$.

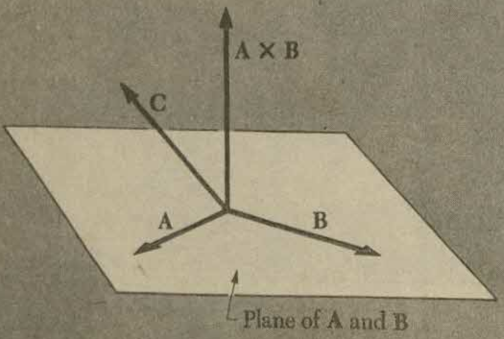
We note from inspection of the figure that

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = (\mathbf{A} \times \mathbf{B}) \cdot \mathbf{C}, \tag{51}$$

so that *the dot and the cross in the scalar triple product may be interchanged without altering the value of the product.* However,

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = -\mathbf{A} \cdot (\mathbf{C} \times \mathbf{B}). \tag{52}$$

A scalar triple product is not altered by permuting cyclically the order of the vectors, but it is reversed in sign if the cyclic order is changed. Cyclic orderings of *ABC* are *BCA* and *CAB;* noncyclic orderings of *ABC* are *BAC* and *ACB.*



A, **B**, and **C** are three vectors. $\mathbf{A} \times \mathbf{B}$ is perpendicular to the plane of **A** and **B**.

3. Law of sines. Consider the triangle defined by $\mathbf{C} = \mathbf{A} + \mathbf{B}$, and take the vector product of both sides of the equation with **A**:

$$\mathbf{A} \times \mathbf{C} = \mathbf{A} \times \mathbf{A} + \mathbf{A} \times \mathbf{B}. \tag{53}$$

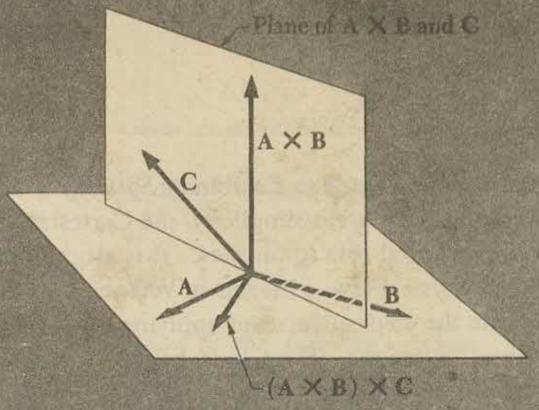
Now $\mathbf{A} \times \mathbf{A} = 0$, and the magnitudes of both sides of (53) must be equal, so that

$$AC \sin|(\mathbf{A},\mathbf{C})| = AB\,|\sin(\mathbf{A},\mathbf{B})|, \tag{54}$$

or

$$\frac{\sin(\mathbf{A},\mathbf{C})}{B} = \frac{\sin(\mathbf{A},\mathbf{B})}{C}, \tag{55}$$

which is the law of sines of a triangle.



(A × B) × C is perpendicular to the plane of A × B and C; it lies in the plane of A and B.

4. Triple products of vectors. There are two types of triple products which are vectors. The vector

$$(\mathbf{A} \cdot \mathbf{B})\,\mathbf{C} \tag{56}$$

is simply the product of the vector **C** by the scalar **A · B**. The other triple product is called the *vector triple product* and is

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}), \tag{57}$$

a vector perpendicular to **A** and to **B × C**. Now **B × C** is perpendicular to the plane of **B** and **C**, so that **A × (B × C)** must lie in the plane of **B** and **C**. By a similar argument

$$(\mathbf{A} \times \mathbf{B}) \times \mathbf{C} = \mathbf{C} \times (\mathbf{B} \times \mathbf{A}) \tag{58}$$

is a vector lying in the plane of **A** and **B** and perpendicular to **C**. It is clear that (57) and (58) are different quantities—the position of the parentheses is critical.

The vector triple product may be expressed as the sum of two terms:

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}\,(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}\,(\mathbf{A} \cdot \mathbf{B}). \tag{59}$$

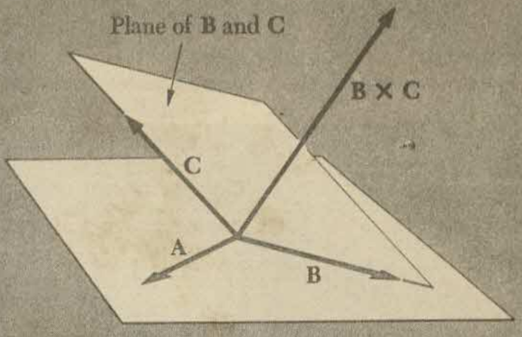
Proofs of this useful relation are given in the standard texts on vector analysis.

5. Force on a particle in a magnetic field. The force on a point electric charge in a magnetic field **B** is proportional to the component of **B** perpendicular to the velocity **v** of the charge. This relation is simply expressed in terms of the vector product:

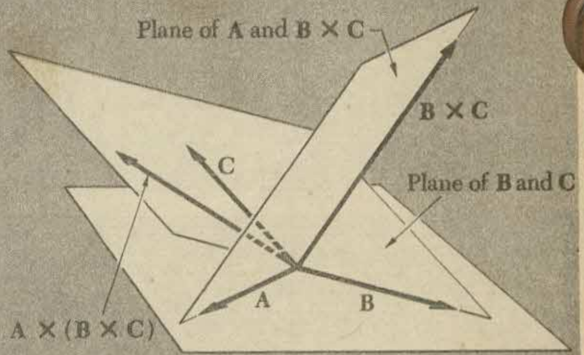
$$\mathbf{F} = \frac{q}{c}\,\mathbf{v} \times \mathbf{B} \qquad \text{(Gaussian units);} \tag{60}$$

$$\mathbf{F} = q\,\mathbf{v} \times \mathbf{B} \qquad \text{(MKS units).} \tag{61}$$

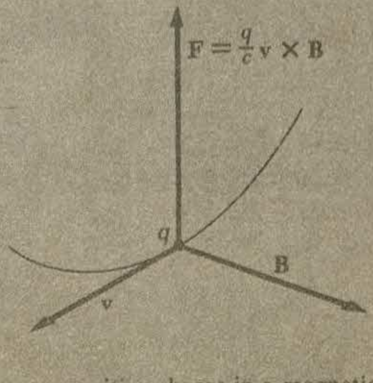
Here q is the charge on the particle and c is the speed of light. This force law is developed in detail in Vol. II.



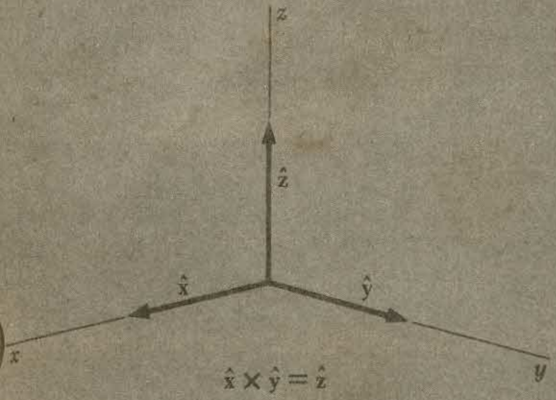
B × C is perpendicular to the plane of B and C.



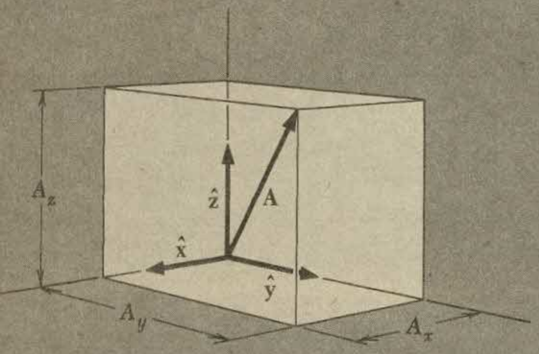
A × (B × C) is perpendicular to the plane of A and (B × C); it lies in the plane of B and C. Obviously A × (B × C) and (A × B) × C are different vectors.



Force on positive charge in a magnetic field.



Cartesian orthogonal unit vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$.



$\mathbf{A} = \hat{\mathbf{x}}A_x + \hat{\mathbf{y}}A_y + \hat{\mathbf{z}}A_z.$

Vectors in the Cartesian Coordinate System

Because of its great simplicity, the Cartesian coordinate system has a privileged role in physics. It is elegant and economical to discuss physical laws in terms of vectors, but ultimately we must evaluate the consequences of applying the laws to very particular physical situations. Here it is helpful to represent a vector in particular coordinate systems, the most useful of which is the Cartesian. This is defined in terms of any three mutually perpendicular unit vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$. There is no misunderstanding if we write these on the blackboard without indicating by a wavy line that they are vectors. Some will prefer to write $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, $\hat{\mathbf{k}}$ for the unit vectors. The direction of $\hat{\mathbf{z}}$ relative to $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ is defined by the right-hand-thread rule as discussed earlier in connection with the vector product. That rule tells us that

$$\hat{\mathbf{z}} = \hat{\mathbf{x}} \times \hat{\mathbf{y}}. \tag{62}$$

We agree to work only in the right-hand coordinate system. How would we communicate our definition of right-handed to a creature in another solar system in our galaxy? We can do this by using circularly polarized radio waves. The signal carries a message which tells the remote observer in which sense we defined the waves to be polarized. The remote observer will have constructed two receivers, one with the correct sense and one with the incorrect, in terms of signal strength. Any method requires clear instructions: In the original analysis of the spectroscopic Zeeman effect its discoverer incorrectly associated a positive sign to the oscillating charges in atoms because he misunderstood the sense of circularly polarized radiation.[†]

Any vector **A** can be written as

$$\mathbf{A} = \hat{\mathbf{x}}A_x + \hat{\mathbf{y}}A_y + \hat{\mathbf{z}}A_z. \tag{63}$$

Here A_x, A_y, A_z are the projections of **A** onto the appropriate coordinate axes; thus

$$A_x = \mathbf{A}\cdot\hat{\mathbf{x}} = A\cos(\mathbf{A},\hat{\mathbf{x}}), \tag{64}$$

and similarly for A_y and A_z. We shall sometimes write **A** as the triplet of numbers (A_x, A_y, A_z). On taking the square of **A** we have

---

[†] See P. Zeeman, *Philosophical Magazine* ser. 5, 43, 55 and 226 (1897). In a similar connection the first Telstar transmission on July 11, 1962, was poorly received in Great Britain because "of the reversal of a small component in the aerial feed which arose from an ambiguity in the accepted definition of the sense of rotation of radio waves." *Times (London)*, July 13, 1962, p. 11.

$$A^2 = A_x^2 + A_y^2 + A_z^2. \tag{65}$$

Invariance of A^2 under Rotation of the Reference Frame. Suppose we leave the direction of the vector **A** unchanged and rotate the reference frame with the coordinate axes about the origin to form a new orthogonal set of unit vectors $\hat{\mathbf{x}}', \hat{\mathbf{y}}', \hat{\mathbf{z}}'$. In the new coordinate system the components of **A** will be $A_{x'}$; $A_{y'}$; $A_{z'}$:

$$\mathbf{A} = \hat{\mathbf{x}}'A_{x'} + \hat{\mathbf{y}}'A_{y'} + \hat{\mathbf{z}}'A_{z'}. \tag{66}$$

There is no problem in thinking of leaving the direction of **A** unchanged—we leave **A** fixed with respect to the fixed stars or, more conveniently, with respect to the printed page. The new reference frame is rotated with respect to the old. The length of A must be independent of the orientation of the reference frame; thus the value of A^2 formed from (66) must be identical with the value of A^2 in (65):

$$A_{x'}^2 + A_{y'}^2 + A_{z'}^2 = A_x^2 + A_y^2 + A_z^2. \tag{67}$$

This is our first example of a *form invariant.*† The form of the magnitude of a vector is the same in all Cartesian coordinate systems which differ by a rigid rotation of the coordinate axes. It is evident from its geometrical definition as a projection that the scalar product

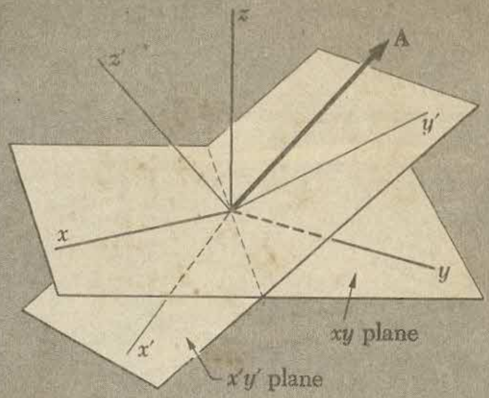
$$\mathbf{A}\cdot\mathbf{B} = A_xB_x + A_yB_y + A_zB_z \tag{68}$$

is another example of a form invariant.

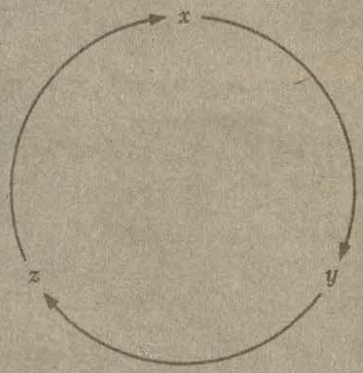
By virtue of its geometrical definition the vector product provides another form invariant. The vector product written in the Cartesian coordinate representation is

$$\begin{aligned}\mathbf{A}\times\mathbf{B} &= (\hat{\mathbf{x}}A_x + \hat{\mathbf{y}}A_y + \hat{\mathbf{z}}A_z)\times(\hat{\mathbf{x}}B_x + \hat{\mathbf{y}}B_y + \hat{\mathbf{z}}B_z)\\ &= (\hat{\mathbf{x}}\times\hat{\mathbf{y}})A_xB_y + (\hat{\mathbf{x}}\times\hat{\mathbf{z}})A_xB_z + (\hat{\mathbf{y}}\times\hat{\mathbf{x}})A_yB_x\\ &\quad + (\hat{\mathbf{y}}\times\hat{\mathbf{z}})A_yB_z + (\hat{\mathbf{z}}\times\hat{\mathbf{x}})A_zB_x + (\hat{\mathbf{z}}\times\hat{\mathbf{y}})A_zB_y,\end{aligned} \tag{69}$$

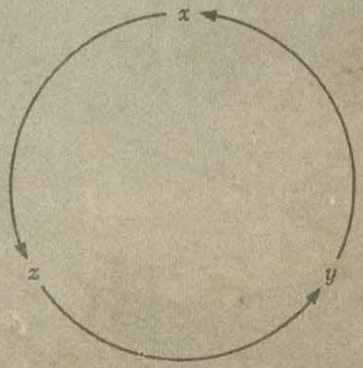
where we have used the relation $\hat{\mathbf{x}}\times\hat{\mathbf{x}} = 0$, etc. Now by the



The vector **A** can be described in coordinates *xyz* or coordinates *x′y′z′* obtained from *xyz* by an arbitrary rotation. We say A^2 is form invariant with respect to the rotation. This means $A_x^2 + A_y^2 + A_z^2 = A_{x'}^2 + A_{y'}^2 + A_{z'}^2$.



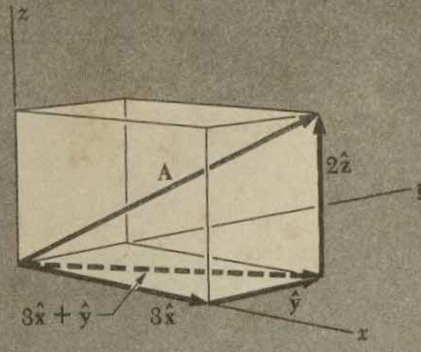
Cyclic order of *xyz*.



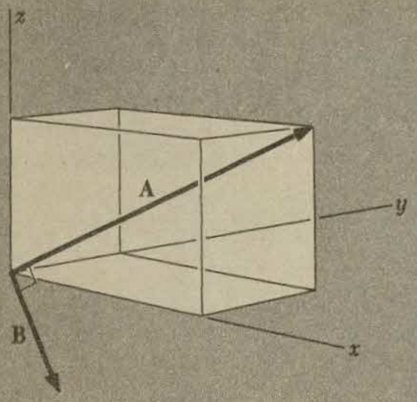
Anticyclic order of *xyz*.

---

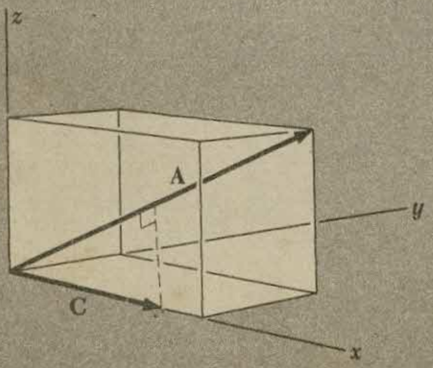
† P. A. M. Dirac: "The important things in the world appear as the invariants (or more generally the nearly invariants, or quantities with simple transformation properties) of these transformations." *Principles of Quantum Mechanics,* 1st ed. (Clarendon Press, Oxford, 1930).



The vector $\mathbf{A} = 3\hat{\mathbf{x}} + \hat{\mathbf{y}} + 2\hat{\mathbf{z}}$ and its projection on the xy plane.



The vector **B** is in the xy plane and perpendicular to **A**.



Projection on **A** of the vector $\mathbf{C} = 2\hat{\mathbf{x}}$.

definition of the vector product we know that

$$\hat{\mathbf{x}} \times \hat{\mathbf{y}} = \hat{\mathbf{z}}; \qquad \hat{\mathbf{y}} \times \hat{\mathbf{z}} = \hat{\mathbf{x}}; \qquad \hat{\mathbf{z}} \times \hat{\mathbf{x}} = \hat{\mathbf{y}}. \tag{70}$$

Observe that in each product the three vectors are in the order xyz and cyclic permutations thereof. The sign changes if the order of the factors is reversed, because the reversal produces an anticyclic ordering of xyz:

$$\hat{\mathbf{y}} \times \hat{\mathbf{x}} = -\hat{\mathbf{z}}; \qquad \hat{\mathbf{z}} \times \hat{\mathbf{y}} = -\hat{\mathbf{x}}; \qquad \hat{\mathbf{x}} \times \hat{\mathbf{z}} = -\hat{\mathbf{y}}. \tag{71}$$

With these results, (69) reduces to

$$\mathbf{A} \times \mathbf{B} = \hat{\mathbf{x}}\,(A_yB_z - A_zB_y) + \hat{\mathbf{y}}\,(A_zB_x - A_xB_z) + \hat{\mathbf{z}}\,(A_xB_y - A_yB_x). \tag{72}$$

Note that if the indices are cyclic with xyz the term enters the vector product with a positive sign; otherwise the sign is negative. If you are familiar with determinants, you can confirm readily that the representation

$$\mathbf{A} \times \mathbf{B} = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ A_x & A_y & A_z \\ B_x & B_y & B_z \end{vmatrix} \tag{73}$$

is equivalent to Eq. (72), and it is easier to remember.

Examples. Various elementary vector operations. We consider the vector

$$\mathbf{A} = 3\hat{\mathbf{x}} + \hat{\mathbf{y}} + 2\hat{\mathbf{z}}.$$

(*a*) Find the length of **A**. We form A^2:

$$A^2 = \mathbf{A} \cdot \mathbf{A} = 3^2 + 1^2 + 2^2 = 14,$$

so that $A = \sqrt{14}$ is the length of **A**.

(*b*) What is the length of the projection of **A** on the xy plane? The vector which is the projection of **A** on the xy plane is $3\hat{\mathbf{x}} + \hat{\mathbf{y}}$; the square of the length of this vector is $3^2 + 1^2 = 10$.

(*c*) Construct a vector in the xy plane and perpendicular to **A**. We want a vector of the form

$$\mathbf{B} = B_x\hat{\mathbf{x}} + B_y\hat{\mathbf{y}}$$

with the property $\mathbf{A} \cdot \mathbf{B} = 0$, or

$$(3\hat{\mathbf{x}} + \hat{\mathbf{y}} + 2\hat{\mathbf{z}}) \cdot (B_x\hat{\mathbf{x}} + B_y\hat{\mathbf{y}}) = 0.$$

On taking the scalar product we find

$$3B_x + B_y = 0,$$

or

$$\frac{B_y}{B_x} = -3.$$

The length of the vector **B** is not determined by the specification of the problem.

(*d*) Construct the unit vector $\hat{\mathbf{B}}$. We must have $\hat{B}_x{}^2 + \hat{B}_y{}^2 = 1$, or

$$\hat{B}_x{}^2\,(1^2 + 3^2) = 1 = 10\,\hat{B}_x{}^2.$$

Thus

$$\hat{\mathbf{B}} = \sqrt{1/10}\,\hat{\mathbf{x}} - \sqrt{9/10}\,\hat{\mathbf{y}} = (\hat{\mathbf{x}} - 3\hat{\mathbf{y}})/\sqrt{10}.$$

(*e*) Find the scalar product with **A** of the vector $\mathbf{C} = 2\hat{\mathbf{x}}$. This is seen directly to be $(2)(3) = 6$.

(*f*) Find the form of **A** and **C** in a reference frame obtained from the old reference frame by a rotation of $\pi/2$ clockwise looking along the positive z axis. The new unit vectors $\hat{\mathbf{x}}'$, $\hat{\mathbf{y}}'$, $\hat{\mathbf{z}}'$ are related to the old $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$ by

$$\hat{\mathbf{x}}' = \hat{\mathbf{y}}; \qquad \hat{\mathbf{y}}' = -\hat{\mathbf{x}}; \qquad \hat{\mathbf{z}}' = \hat{\mathbf{z}}.$$

Where $\hat{\mathbf{x}}$ appeared, we now have $-\hat{\mathbf{y}}'$; where $\hat{\mathbf{y}}$ appeared, we now have $\hat{\mathbf{x}}'$, so that

$$\mathbf{A} = \hat{\mathbf{x}}' - 3\hat{\mathbf{y}}' + 2\hat{\mathbf{z}}'; \qquad \mathbf{C} = -2\hat{\mathbf{y}}'.$$

(*g*) Find the scalar product $\mathbf{A} \cdot \mathbf{C}$ in the primed coordinate system. From the result of (*f*) we get $(-3)(-2) = 6$, exactly as in the unprimed system.

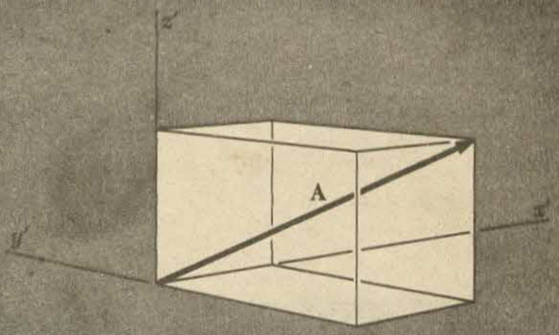
(*h*) Find the vector product $\mathbf{A} \times \mathbf{C}$. In the unprimed system that is

$$\begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ 3 & 1 & 2 \\ 2 & 0 & 0 \end{vmatrix} = 4\hat{\mathbf{y}} - 2\hat{\mathbf{z}}.$$

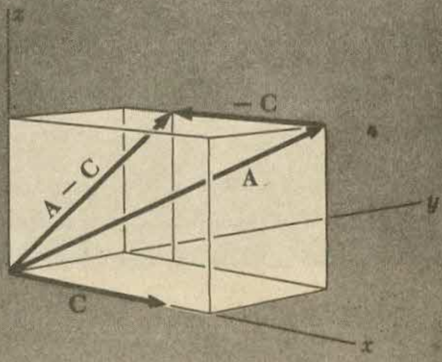
By forming the scalar products, you may confirm that this vector is perpendicular both to **A** and to **C**.

(*i*) Form the vector $\mathbf{A} - \mathbf{C}$. We have

$$\mathbf{A} - \mathbf{C} = (3 - 2)\hat{\mathbf{x}} + \hat{\mathbf{y}} + 2\hat{\mathbf{z}} = \hat{\mathbf{x}} + \hat{\mathbf{y}} + 2\hat{\mathbf{z}}.$$



The primed reference frame x', y', z' is generated from the unprimed system x, y, z, by a rotation of $+\pi/2$ about the z axis.



Vector A − C.

The vector product has a particular utility in describing the angular velocity and angular acceleration of a rotating body. We have seen that finite rotations are not vectors because two such rotations do not obey the vector law of addition. But the angular velocity is defined as the limit of an infinitesimal rotation angle divided by the infinitesimal time interval in which the rotation occurs. The order in which two infinitesimal rotations is performed does not affect the final position of an object, except for terms that are as small as the square of the magnitude of the infinitesimal rotations, and this vanishes in the appropriate limit. We prove this result and discuss the elementary physics of rotating bodies in a later chapter.

We sometimes speak of a scalar function of position, such as the temperature $T(x,y,z)$ at the point (x,y,z), as a *scalar field*. Similarly, a vector whose value is a function of position, such as the velocity $\mathbf{v}(x,y,z)$, of a particle at the point (x,y,z) is spoken of as a *vector field*. Much of the subject of vector analysis is concerned with scalar and vector fields and with differential operations on vectors, which are discussed fully in Vol. II.

Useful vector identities

$$\mathbf{A}\cdot\mathbf{B} = A_xB_x + A_yB_y + A_zB_z; \tag{74}$$

$$\begin{aligned}\mathbf{A}\times\mathbf{B} = \hat{\mathbf{x}}\,(A_yB_z - A_zB_y) + \hat{\mathbf{y}}\,(A_zB_x - A_xB_z)\\ + \hat{\mathbf{z}}\,(A_xB_y - A_yB_x);\end{aligned} \tag{75}$$

$$(\mathbf{A}\times\mathbf{B})\times\mathbf{C} = (\mathbf{A}\cdot\mathbf{C})\mathbf{B} - (\mathbf{B}\cdot\mathbf{C})\mathbf{A}; \tag{76}$$

$$\mathbf{A}\times(\mathbf{B}\times\mathbf{C}) = (\mathbf{A}\cdot\mathbf{C})\,\mathbf{B} - (\mathbf{A}\cdot\mathbf{B})\,\mathbf{C}; \tag{77}$$

$$(\mathbf{A}\times\mathbf{B})\cdot(\mathbf{C}\times\mathbf{D}) = (\mathbf{A}\cdot\mathbf{C})(\mathbf{B}\cdot\mathbf{D}) - (\mathbf{A}\cdot\mathbf{D})(\mathbf{B}\cdot\mathbf{C}); \tag{78}$$

$$\begin{aligned}(\mathbf{A}\times\mathbf{B})\times(\mathbf{C}\times\mathbf{D}) = [\mathbf{A}\cdot(\mathbf{B}\times\mathbf{D})]\,\mathbf{C}\\ - [\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C})]\,\mathbf{D};\end{aligned} \tag{79}$$

$$\mathbf{A}\times[\mathbf{B}\times(\mathbf{C}\times\mathbf{D})] = (\mathbf{A}\times\mathbf{C})(\mathbf{B}\cdot\mathbf{D}) - (\mathbf{A}\times\mathbf{D})(\mathbf{B}\cdot\mathbf{C}). \tag{80}$$

Further reading

PSSC, chap. 6.

G. E. Hay, *Vector and tensor analysis* (Dover, New York, 1953).

D. E. Rutherford, *Vector methods* (Oliver and Boyd, Edinburgh; or Interscience, New York; 1949).

C. E. Weatherburn, *Elementary and advanced vector analysis*, 2 vols. (G. Bell, London, 1928).

Film list

"Vector Kinematics" (16 min) F. Friedman (PSSC-MLA 0109). A computer traces out on a cathode-ray tube the velocity and acceleration vectors corresponding to various types of displacements of a spot: circular, simple harmonic, and free fall.

"Electrons in a Uniform Magnetic Field" D. Montgomery (PSSC-MLA 0412). Shows the behavior of the electron beam in the Leybold e/m tube.

Problems

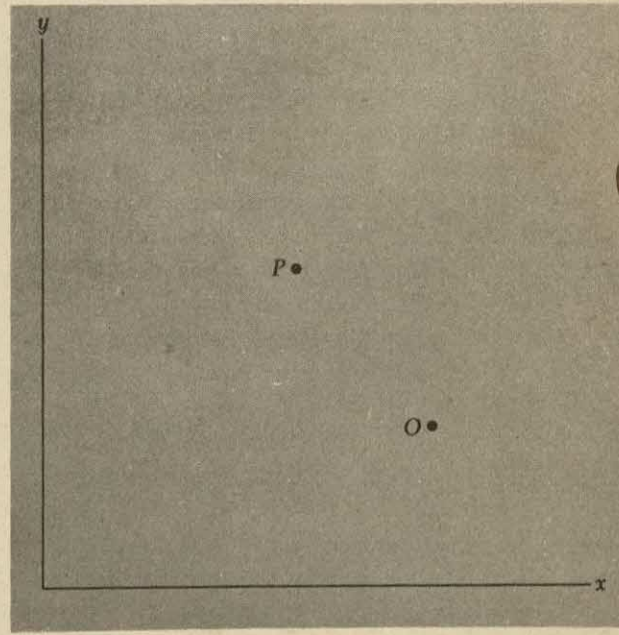
In these problems use vector methods wherever possible:

1. *Components of a vector.* On a piece of paper place two points, labeling them O and P. Draw a line from O to P, placing an arrowhead at the end.

(*a*) Find the x and y components of the vector, in centimeters. (1 in. = 2.54 cm.)

(*b*) Parallel to these axes draw another set passing through the point O; what are the new x' and y' components?

(*c*) Rotate the second set of axes 30° counterclockwise (as you look at the paper) and find the x'' and y'' components.



2. *Addition of vectors.* Draw the result of the following vector additions:

(*a*) Add a vector 2 cm east to one 3 cm northwest.

(*b*) Add a vector 8 cm east to one 12 cm northwest.

(*c*) Compare the results of parts (*a*) and (*b*), and frame a theorem about adding a pair of vectors that are multiples of another pair.

3. *Multiplication by a scalar.* Let $\mathbf{A} = 2.0$ cm at 70° east of north and $\mathbf{B} = 3.5$ cm at 130° east of north. Use either a protractor or polar coordinate graph paper in your solutions.

(*a*) Draw the vectors described above and two others 2.5 times as large.

(*b*) Multiply **A** by -2 and **B** by $+3$ and find the vector sum.

Ans. 9.4 cm at 150°.

(*c*) Place a point 10 cm due north of the origin. Find multiples of **A** and **B** whose vector sum is the vector from the origin to this point.

4. *Unit vectors.* (*a*) Draw a vector of unit length, multiply it by 4, and draw the new vector.

(*b*) Draw a second unit vector at right angles to the first vector, the one of length 4. Multiply this by -3 and add this to the new vector of part (*a*).

(*c*) Using coordinate axes oriented along the unit vectors, find the x and y components of the sum vector.

(*d*) An arbitrary vector can be expressed as the vector sum of certain multiples of any two vectors. What is the particular utility of orthogonal (perpendicular) unit vectors? (*Hint:* Consider the scalar product of two vectors.)

5. *Scalar and vector products of two vectors.* Given two vectors $\mathbf{a} = 3\hat{x} + 4\hat{y} - 5\hat{z}$ and $\mathbf{b} = -\hat{x} + 2\hat{y} + 6\hat{z}$, calculate by vector methods:

(*a*) The length of each. *Ans.* $a = \sqrt{50}$; $b = \sqrt{41}$.

(*b*) The scalar product $\mathbf{a} \cdot \mathbf{b}$. *Ans.* -25.

(*c*) The included angle between them. *Ans.* 123.5°.

(*d*) The direction cosines for each.

(*e*) The vector sum and difference $\mathbf{a} + \mathbf{b}$ and $\mathbf{a} - \mathbf{b}$.

Ans. $\mathbf{a} + \mathbf{b} = 2\hat{x} + 6\hat{y} + \hat{z}$.

(*f*) The vector product $\mathbf{a} \times \mathbf{b}$. *Ans.* $34\hat{x} - 13\hat{y} + 10\hat{z}$.

6. *Vector algebra.* Given two vectors such that $\mathbf{a} + \mathbf{b} = 11\hat{x} - \hat{y} + 5\hat{z}$ and $\mathbf{a} - \mathbf{b} = -5\hat{x} + 11\hat{y} + 9\hat{z}$:

(*a*) Find $\mathbf{a}$ and $\mathbf{b}$.

(*b*) Find the angle included between $\mathbf{a}$ and $(\mathbf{a} + \mathbf{b})$, using vector methods.

7. *Composition of velocities.* The pilot of an airplane wishes to reach a point 200 miles east of his present position. A wind blows 30 miles per hour from the northwest. Calculate his vector velocity with respect to the moving air mass if his schedule requires him to arrive at his destination in 40 minutes.

Ans. $\mathbf{v} = 279\hat{x} + 21\hat{y}$ miles/hour; $\hat{x}$ = east; $\hat{y}$ = north.

8. *Inversion of coordinate axes.* It is possible to turn a right-handed unit coordinate system into a left-handed unit system by multiplying all three unit vectors by one scalar. What is this number?

9. *Vector operations; relative position vector.* Two particles are emitted from a common source and at a particular time have displacements:

$$\mathbf{r}_1 = 4\hat{x} + 3\hat{y} + 8\hat{z}; \qquad \mathbf{r}_2 = 2\hat{x} + 10\hat{y} + 5\hat{z}.$$

(*a*) Sketch the positions of the particles and write the expression for the displacement $\mathbf{r}$ of particle 2 relative to particle 1.

(*b*) Use the scalar product to find the magnitude of each vector.

Ans. $r_1 = 9.4$; $r_2 = 11.3$; $r = 7.9$.

(*c*) Calculate the angles between all possible pairs of the three vectors.

(*d*) Calculate the projection of $\mathbf{r}$ on $\mathbf{r}_1$. *Ans.* -1.2.

(*e*) Calculate the vector product $\mathbf{r}_1 \times \mathbf{r}_2$. *Ans.* $-65\hat{x} - 4\hat{y} + 34\hat{z}$.

10. *Closest approach of two particles.* Two particles 1 and 2 travel along the x and y axes with respective velocities $\mathbf{v}_1 = 2\hat{x}$ cm/sec and $\mathbf{v}_2 = 3\hat{y}$ cm/sec. At $t = 0$ they are at

$$x_1 = -3 \text{ cm}, \quad y_1 = 0; \qquad x_2 = 0, \quad y_2 = -3 \text{ cm}.$$

(*a*) Find the vector $\mathbf{r}_2 - \mathbf{r}_1$ which represents the position of 2 relative to 1, as a function of time. *Ans.* $\mathbf{r} = (3 - 2t)\hat{x} + (3t - 3)\hat{y}$ cm.

(*b*) When and where are these two particles closest? *Ans.* $t = 1.15$ sec.

11. *Body diagonals of a cube.* What is the angle between two intersecting body diagonals of a cube? (A *body diagonal* connects two corners and passes through the interior of the cube. A *face diagonal* connects two corners and runs on one face of the cube.)

12. *Condition for* $\mathbf{a} \perp \mathbf{b}$. Show that $\mathbf{a}$ is perpendicular to $\mathbf{b}$ if $|\mathbf{a} + \mathbf{b}| = |\mathbf{a} - \mathbf{b}|$.

13. *Area of tetrahedron.* Consider a tetrahedron whose vertices O, A, B, C lie respectively at the origin and on the x, y, z axes. (The position vector of A is $\mathbf{a} = a\,\hat{\mathbf{x}}$, etc.) Develop an expression for its total surface area, which will be the appropriate sum of the areas of the triangular faces.

14. *Volume of parallelepiped.* A parallelepiped has edges described by the vectors $\hat{\mathbf{x}} + 2\hat{\mathbf{y}}$; $4\hat{\mathbf{y}}$; and $\hat{\mathbf{y}} + 3\hat{\mathbf{z}}$ from the origin. Find the volume.

Ans. 12.

15. *Equilibrium of forces.* Three forces $\mathbf{F}_1$, $\mathbf{F}_2$, and $\mathbf{F}_3$ act simultaneously on a point particle. The resultant force $\mathbf{F}_R$ is simply the vector sum of the forces. The particle is said to be in equilibrium if $\mathbf{F}_R = 0$.

(*a*) Show that if $\mathbf{F}_R = 0$ the vectors representing three forces form a triangle.

(*b*) If $\mathbf{F}_R = 0$ as above, is it possible for any one of the vectors to lie outside the plane determined by the other two?

(*c*) A particle subject to a vertically downward force of 10 newtons and suspended from a cord (tension 15 newtons) making an angle of 0.1 radian from the vertical cannot be in equilibrium. What third force is required to produce equilibrium? Is the answer unique?

16. *Work done by forces.* The constant forces $\mathbf{F}_1 = \hat{\mathbf{x}} + 2\hat{\mathbf{y}} + 3\hat{\mathbf{z}}$ (dynes) and $\mathbf{F}_2 = 4\hat{\mathbf{x}} - 5\hat{\mathbf{y}} - 2\hat{\mathbf{z}}$ (dynes) act together on a particle, during a displacement from the point $A(20,15,0)$ (cm) to the point $B(0,0,7)$ (cm).

(*a*) What is the work done (in ergs) on the particle? The work done (Chap. 5) is given by $\mathbf{F} \cdot \mathbf{r}$, where $\mathbf{F}$ is the resultant force (here $\mathbf{F} = \mathbf{F}_1 + \mathbf{F}_2$) and $\mathbf{r}$ is the displacement. *Ans.* -48 ergs.

(*b*) Suppose the same forces were acting, but the motion went from B to A. What is the work done on the particle in this situation?

17. *Torque of force about a point.* The torque or turning moment $\mathbf{N}$ of a force about a given point is given by $\mathbf{r} \times \mathbf{F}$, where $\mathbf{r}$ is the vector from the given point to the point of application of $\mathbf{F}$. Consider a force $\mathbf{F} = -3\hat{\mathbf{x}} + \hat{\mathbf{y}} + 5\hat{\mathbf{z}}$ (dynes) acting at the point $7\hat{\mathbf{x}} + 3\hat{\mathbf{y}} + \hat{\mathbf{z}}$ (cm). Remember that $\mathbf{F} \times \mathbf{r} = -\mathbf{r} \times \mathbf{F}$.

(*a*) What is the torque in dyne-cm about the origin? (Just give the result for $\mathbf{N}$ as a linear combination of $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$.)

Ans. $14\hat{\mathbf{x}} - 38\hat{\mathbf{y}} + 16\hat{\mathbf{z}}$.

(*b*) What is the torque about the point (0,10,0)?

Ans. $36\hat{\mathbf{x}} - 38\hat{\mathbf{y}} - 14\hat{\mathbf{z}}$.

18. *Vector product as a determinant.* Note that the expression for an expansion of $\mathbf{A} \times \mathbf{B}$ in terms of components involves terms of the form $A_iB_j - B_iA_j$. This is just the expansion of the second-order determinant

$$\begin{vmatrix} A_i & A_j \\ B_i & B_j \end{vmatrix}.$$

(*a*) Show that, if the unit vectors are handled as numbers,

$$\mathbf{A} \times \mathbf{B} = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ A_x & A_y & A_z \\ B_x & B_y & B_z \end{vmatrix}.$$

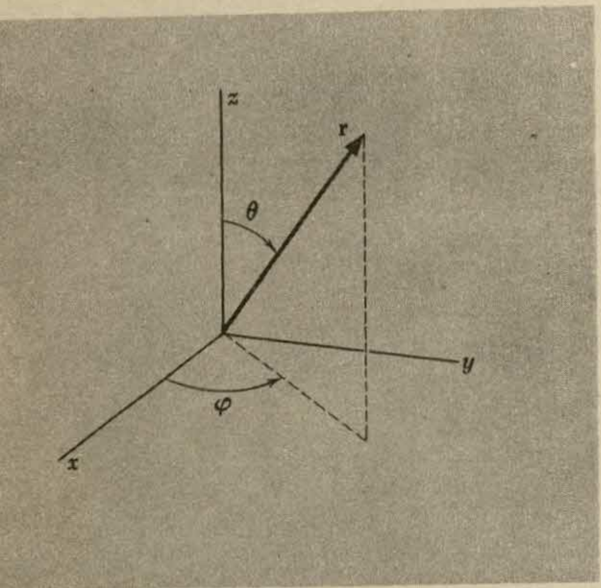
(*b*) Show further that

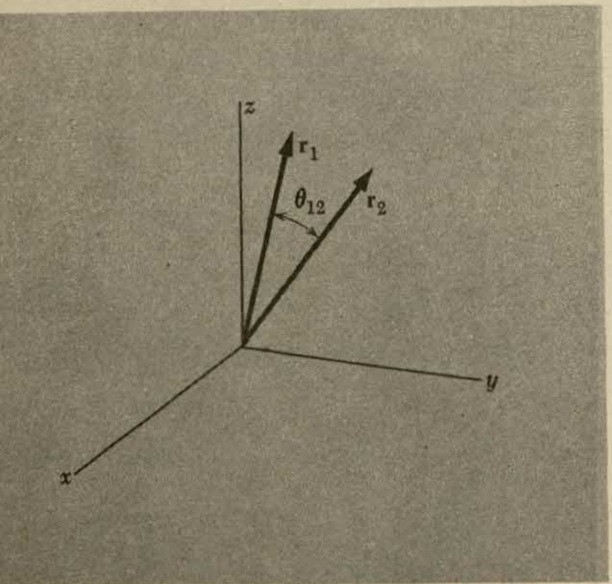
$$\mathbf{C}\cdot(\mathbf{A}\times\mathbf{B}) = \begin{vmatrix} C_x & C_y & C_z \\ A_x & A_y & A_z \\ B_x & B_y & B_z \end{vmatrix}.$$

19. *Random flights.* A particle follows in space a path which consists of N equal steps, each of length s. The direction in space of each step is entirely random, with no relation or correlation between any two steps. The total displacement is

$$\mathbf{S} = \sum_{i=1}^{N} \mathbf{s}_i.$$

Show that the mean square displacement between initial and final positions is $\langle S^2 \rangle = Ns^2$, where $\langle\ \rangle$ denotes mean value. [*Hint:* The assumption that the direction of every step is independent of the direction of every other step means that $\langle \mathbf{s}_i \cdot \mathbf{s}_j \rangle = 0$ for all i and j, except $i = j$.]





Advanced Topic 1. Vectors and spherical polar coordinates

The position of a particle is expressed in spherical polar coordinates as r, θ, φ. Here r is the magnitude of the vector $\mathbf{r}$ from the origin to the particle; θ is the angle between $\mathbf{r}$ and the polar axis z; and φ is the angle between the x axis and the projection of r on the equatorial or xy plane. We take $0 \leqq \theta \leqq \pi$. The projection of $\mathbf{r}$ on the xy plane is of magnitude $r \sin\theta$. Note that the position in Cartesian coordinates is given by

$$x = r\sin\theta\cos\varphi; \qquad y = r\sin\theta\sin\varphi; \qquad z = r\cos\theta. \tag{81}$$

(*a*) Let the first particle be at $\mathbf{r}_1 \equiv (r_1, \theta_1, \varphi_1)$ and a second particle at $\mathbf{r}_2 \equiv (r_2, \theta_2, \varphi_2)$. Let θ_{12} be the angle between $\mathbf{r}_1$ and $\mathbf{r}_2$. By expressing the scalar product $\hat{\mathbf{r}}_1 \cdot \hat{\mathbf{r}}_2 = \cos\theta_{12}$ in terms of $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$, show that

$$\cos\theta_{12} = \sin\theta_1\sin\theta_2\cos(\varphi_1 - \varphi_2) + \cos\theta_1\cos\theta_2, \tag{82}$$

where we have used the trigonometric identity

$$\cos(\varphi_1 - \varphi_2) = \cos\varphi_1\cos\varphi_2 + \sin\varphi_1\sin\varphi_2. \tag{83}$$

This is a good example of the power of vector methods. [Try to find the result (82) otherwise!]

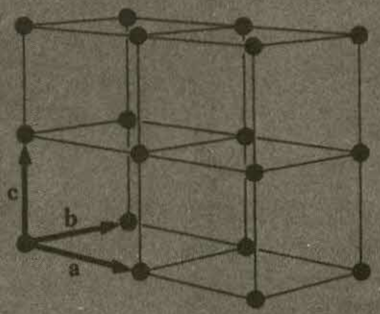
(*b*) Similarly, by forming the vector product $\hat{\mathbf{r}}_1 \times \hat{\mathbf{r}}_2$, find a relation for $\sin\theta_{12}$.

Cylindrical polar coordinates ρ, φ, z are an orthogonal set of coordinates defined by $x = \rho\cos\varphi$; $y = \rho\sin\varphi$; $z = z$. When used in two dimensions, the coordinates reduce to ρ and φ alone.

Advanced Topic 2. Crystal lattices and the reciprocal lattice

Within a certain domain we have assumed that space is Euclidean. Our assumption means that if *everything* involved in an experiment or an observation is moved along by the same vector displacement (denoted by **t**), nothing is changed in the outcome of the experiment. We say then that the laws of physics are invariant (unchanged) under all translations **t**. In principle we must move everything having any connection with the experiment—it will not do to move a pendulum from sea level to the top of Mt. Everest: We know that such a move relative to the environment will change the frequency of the pendulum because of the changed value of *g*, the acceleration due to gravity.

A crystal is a regular array of atoms in space. If we keep the crystal *fixed* but move about ourselves within the crystal, we will see an environment which changes very considerably from point to point. The world as seen by an electron moving inside a crystal which is held fixed does not have the full translational invariance of the world outside the crystal in free space. But the environment in the crystal about a point **r**′ will be the same as the environment about a point **r** if the two points are separated by a repeat distance of the crystal lattice.



View of a small part of a crystal lattice, with fundamental translation vectors **a**, **b**, **c**. In the particular case shown **a**, **b**, and **c** are orthogonal, but for many crystals **a**, **b**, and **c** are not orthogonal.

We define an ideal crystal as a body composed of atoms arranged in a lattice in such a way that there exist three fundamental translation vectors **a**, **b**, **c** with the property that the atomic arrangement looks the same in every respect when viewed from any point **r** as when viewed from the point

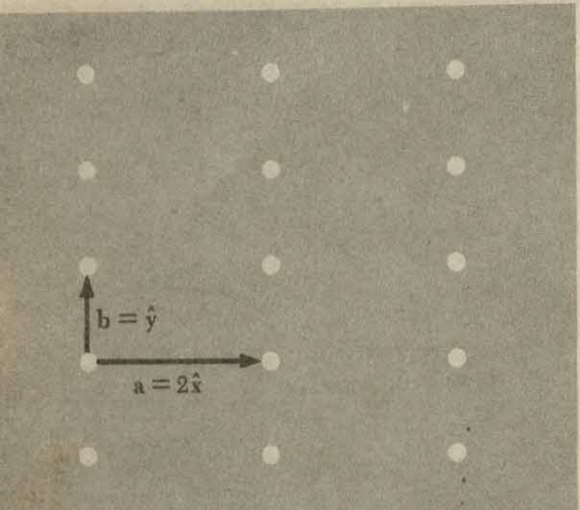
$$\mathbf{r}' = \mathbf{r} + n_1\mathbf{a} + n_2\mathbf{b} + n_3\mathbf{c}, \tag{84}$$

where n_1, n_2, n_3 are arbitrary integers. Note that **a**, **b**, **c** are not vectors of unit length: they are usually of atomic dimensions, or somewhat larger. A set of fundamental translation vectors is said to be *primitive* if any two points **r**, **r**′ from which the atomic arrangement looks the same can be connected by a relation of the form of (84) for a suitable choice of the integers n_1, n_2, n_3.

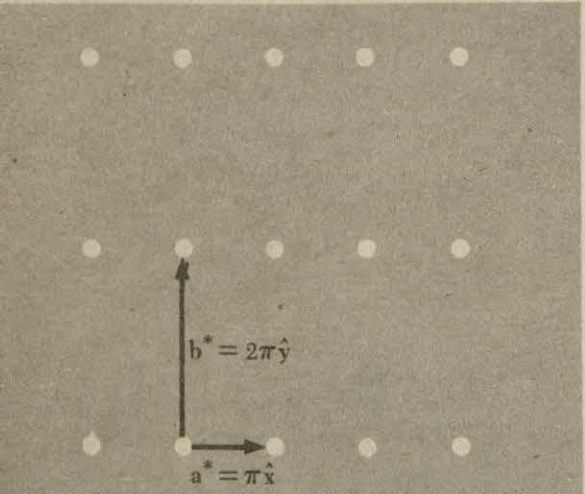
Thus the world inside a fixed crystal is invariant not under just any translation but only under translations of the form

$$\mathbf{t} = n_1\mathbf{a} + n_2\mathbf{b} + n_3\mathbf{c}, \tag{85}$$

where n_1, n_2, n_3 are integers. The laws of mechanics of a mobile electron inside a fixed crystal may be quite unlike the simple laws which prevail in empty space. A discussion which will be fully accessible after you have studied Vol. IV on quantum physics is given in C. Kittel, *Elementary solid state physics*, chap. 6 (Wiley, New York, 1962).



Direct crystal lattice for $a = 2\hat{x}$ and $b = \hat{y}$.



Reciprocal lattice with $a^* = \pi\hat{x}$ and $b^* = 2\pi\hat{y}$. The scale is not the same as in the preceding figure.

More often than not the primitive translation vectors **a**, **b**, **c** are not orthogonal. The mathematical treatment of crystal problems and particularly of the diffraction of x-rays and electrons by crystal lattices is greatly simplified with the help of the reciprocal lattice transformation introduced by J. W. Gibbs. The basis vectors **a***, **b***, **c*** of the reciprocal lattice are defined in terms of the primitive basis vectors **a**, **b**, **c** of the direct lattice by the equations

$$\mathbf{a}^* = 2\pi\frac{\mathbf{b}\times\mathbf{c}}{\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})};\ \mathbf{b}^* = 2\pi\frac{\mathbf{c}\times\mathbf{a}}{\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})};\ \mathbf{c}^* = 2\pi\frac{\mathbf{a}\times\mathbf{b}}{\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})}. \qquad (86)$$

This definition includes a useful factor 2π which is not present in the usual crystallographic definition followed by elementary texts.

Note: The units for vectors in the reciprocal lattice are not units of distance. If **a**, **b**, **c**, are measured in centimeters (cm), then **a***, **b***, **c*** are in reciprocal centimeters (cm^{-1}).

(*a*) Show that $\mathbf{a}^* \cdot \mathbf{a} = 2\pi$ and that $\mathbf{a}^* \cdot \mathbf{b} = 0$.

(*b*) A two-dimensional crystal lattice has, in appropriate units,

$$\mathbf{a} = 4\hat{\mathbf{x}}, \qquad \mathbf{b} = \hat{\mathbf{y}},$$

where $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are orthogonal Cartesian axes. Sketch part of the lattice. Find **a*** and **b***, and sketch part of the reciprocal lattice. *Hint:* Let $\mathbf{c} = \hat{\mathbf{z}}$ if this will make you feel easier about the appropriate definitions in two dimensions.

(*c*) Another two-dimensional lattice has

$$\mathbf{a} = \hat{\mathbf{x}}, \qquad \mathbf{b} = \hat{\mathbf{x}}\cos\frac{\pi}{3} + \hat{\mathbf{y}}\sin\frac{\pi}{3}.$$

Sketch part of the lattice; find the reciprocal lattice vectors and sketch the reciprocal lattice.

(*d*) By slightly tedious vector manipulation show that

$$V^* = \frac{(2\pi)^3}{V},$$

where V is the volume $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})$ of a primitive cell in the direct lattice and V^* is the volume $\mathbf{a}^*\cdot(\mathbf{b}^*\times\mathbf{c}^*)$ of a primitive cell in the reciprocal lattice.

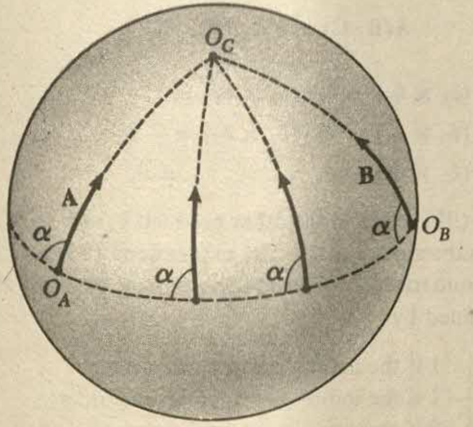
Mathematical Note 1. Equality of vectors in a spherical space

Our concept of the equality of two vectors depends upon the assumption that the space is Euclidean. In a curved space we cannot compare unambiguously two vectors at different points. Consider the two-dimensional curved space formed by the surface of a three-dimensional sphere. We define straight lines in this space as great circles on the sphere, for these are the shortest distances between points.

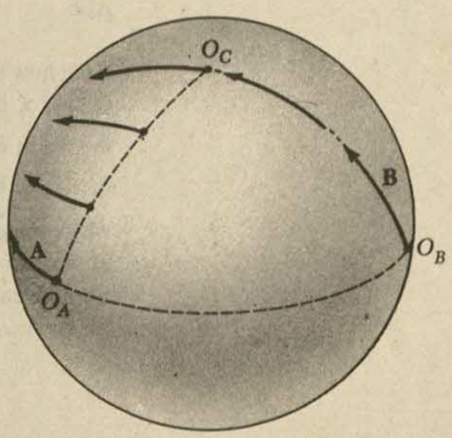
If in this space we wish to compare two vectors **A** and **B**, we move **B** without changing its direction until its initial point coincides with the initial point of **A**. But how do we define "without changing direction"? Following the construction used in flat space, we draw the straight line between O_A and O_B. We then move **B** along toward **A** so that the angle of **B** with the straight line O_AO_B is unchanged.

We can check on our procedure by first moving **B** to O_C to compare it with **C** and then moving **B** along the line O_CO_A to compare it again with **A**. But these two different ways of comparing **B** with **A** give different answers. Suppose, to consider a particular configuration, that **A** and **B** are perpendicular to the line O_AO_B and therefore parallel to each other. Let **C** be chosen to be parallel to **B** in the sense that **C** and **B** make equal angles with the line O_CO_B. But it is perfectly evident from the figure that **C** is not parallel to **A** because they do not make equal angles with the line connecting O_C and O_A. Thus we have a contradiction: **B** is parallel to **A** and **C** is parallel to **B**, but **C** is not parallel to **A**, taking the shortest routes in each comparison. If we take **B** first to **C** and then to **A**, we find now that **B** is not parallel to **A**.

For an exercise suppose that in the second measurement **B** is found to be perpendicular to **A**. How long are the sides of the spherical triangle, expressed in terms of the circumference of the sphere?



One way to compare the direction of **B** with **A** is to move **B** (starting at O_B) along the equator, but always pointing toward O_C, until **B** gets to O_A.



Here, **B** goes up along its meridian until its tail is at O_C—then it slides sideways down to the equator.

Mathematical Note 2. Generalized Cartesian vector notation

The manipulation and evaluation of vector expressions in terms of the standard Cartesian system is facilitated by a slight change in notation. The new notation permits the explicit expression of the Cartesian components of vector products in a more concise form than that developed earlier.

We denote the Cartesian basis vectors by $\hat{\mathbf{e}}_1$, $\hat{\mathbf{e}}_2$, $\hat{\mathbf{e}}_3$ respectively, and define the *Kronecker delta* symbol δ_{ij} as

$$\delta_{ij} \equiv \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases} \tag{87}$$

where i, j run over 1, 2, 3. The expression of a vector $\mathbf{A}$ in terms of its Cartesian components then becomes

$$\mathbf{A} = A_1\hat{\mathbf{e}}_1 + A_2\hat{\mathbf{e}}_2 + A_3\hat{\mathbf{e}}_3 = \sum_{i=1}^{3} A_i\hat{\mathbf{e}}_i. \tag{88}$$

The Greek capital sigma Σ denotes an instruction to sum the quantity following the Σ over the range of values of the indices indicated above and below the Σ. We establish the convention that summation is to be carried out over any *repeated Greek index.* That is,

$$A_\mu e_\mu \equiv \sum_{i=1}^{3} A_i\hat{\mathbf{e}}_i = \mathbf{A}. \tag{89}$$

Then, for example, we have the scalar product

$$\mathbf{A}\cdot\mathbf{B} = \left(\sum_{i=1}^{3} A_i\hat{\mathbf{e}}_i\right)\cdot\left(\sum_{j=1}^{3} B_j\hat{\mathbf{e}}_j\right) = \sum_{i,j=1}^{3} A_iB_j\hat{\mathbf{e}}_i\cdot\hat{\mathbf{e}}_j = \sum_{i=1}^{3} A_iB_i = A_\mu B_\mu, \tag{90}$$

using the relation

$$\hat{\mathbf{e}}_i\cdot\hat{\mathbf{e}}_j = \delta_{ij}. \tag{91}$$

Also

$$\mathbf{A}(\mathbf{B}\cdot\mathbf{C}) = \hat{\mathbf{e}}_\mu A_\mu B_\nu C_\nu. \tag{92}$$

Note now that

$$\hat{\mathbf{e}}_1\cdot(\hat{\mathbf{e}}_2\times\hat{\mathbf{e}}_3) = \hat{\mathbf{e}}_3\cdot(\hat{\mathbf{e}}_1\times\hat{\mathbf{e}}_2) = \hat{\mathbf{e}}_2\cdot(\hat{\mathbf{e}}_3\times\hat{\mathbf{e}}_1) = 1; \tag{93}$$

$$\hat{\mathbf{e}}_1\cdot(\hat{\mathbf{e}}_3\times\hat{\mathbf{e}}_2) = \hat{\mathbf{e}}_3\cdot(\hat{\mathbf{e}}_2\times\hat{\mathbf{e}}_1) = \hat{\mathbf{e}}_2\cdot(\hat{\mathbf{e}}_1\times\hat{\mathbf{e}}_3) = -1; \tag{94}$$

$$\hat{\mathbf{e}}_1\cdot(\hat{\mathbf{e}}_1\times\hat{\mathbf{e}}_2) = \hat{\mathbf{e}}_1\cdot(\hat{\mathbf{e}}_2\times\hat{\mathbf{e}}_1) = \ldots\ldots\ldots\ldots = 0, \tag{95}$$

where the ellipsis in (95) refers to the other products $\hat{\mathbf{e}}_1\cdot(\hat{\mathbf{e}}_j\times\hat{\mathbf{e}}_k)$ having at least two of the subscripts equal. The expressions (93) to (95) may be conveniently summarized by introducing a new symbol, the *Levi-Civita tensor* ϵ_{ijk}, defined by

$$\epsilon_{ijk} \equiv \begin{cases} 1 \text{ if the indices are in cyclic order;} \\ -1 \text{ if the indices are in anticyclic order;} \\ 0 \text{ if any two indices are the same,} \end{cases} \tag{96}$$

where the indices i, j, k range over 1, 2, 3. Here ϵ is the Greek epsilon. *Cyclic order* refers to the orderings 123, 312, 231; *anticyclic* refers to the orderings 132, 321, 213. Equations (93) and (94) take the form

$$\hat{\mathbf{e}}_i \cdot (\hat{\mathbf{e}}_j \times \hat{\mathbf{e}}_k) = \epsilon_{ijk} \qquad (i, j, k = 1, 2, 3). \tag{97}$$

The expression in Cartesian components of the vector product becomes

$$\mathbf{A} \times \mathbf{B} = \sum_{i,j,k=1}^{3} \epsilon_{ijk}\, \hat{\mathbf{e}}_i A_j B_k = \epsilon_{\lambda\mu\nu}\, \hat{\mathbf{e}}_\lambda A_\mu B_\nu. \tag{98}$$

As an exercise, use (98) to evaluate $(\mathbf{A} \times \mathbf{B})^2$.

Historical Note. J. W. Gibbs

An account of the origins of vector analysis is given in Chap. 7 of *Josiah Willard Gibbs: the history of a great mind,* by Lynde Phelps Wheeler (Yale University Press, New Haven, Conn., paperback edition, 1962). Gibbs perhaps carried too far a concern with notation, as when he wrote, "It is really this hesitation in regard to the merits of dif[ferent] notations wh[ich] has kept me from publishing anything on the subject before, and wh[ich] has now kept me from any form of permanent publication . . . I was deterred by a certain feeling of crudeness in my use of symbols." The notation of the dot and the cross for the scalar and vector products is due to Gibbs. The development of vector analysis illustrates an attitude characteristic of Gibbs: "If I have had any success in mathematical physics, it is, I think, because I have been able to dodge mathematical difficulties."

Elegant and clear lectures by Gibbs on vector analysis are published in J. W. Gibbs, *Vector analysis,* edited by E. B. Wilson (Yale University Press, New Haven, 1901).

Chapter 3

Review: Newton's laws of motion 55
Example: Ultracentrifuge 56
Inertial reference frames 58
Forces in inertial reference systems 61
Absolute and relative acceleration 63
Absolute and relative velocity 64
Galilean transformation 66
Conservation of momentum 69
Chemical reactions 73
Example: Collision of heavy particle with light particle 74
Fictitious forces 75
Examples: Accelerometer 76
Centrifugal force 76
Freely falling elevator 77
Foucault pendulum 77
Newtonian law of universal gravitation 79
Example: Satellite in circular orbit 80

Further reading 81

Film list 81

Problems 81

Advanced Topic: Velocity and acceleration in rotating coordinate systems 83
Examples: Uniform linear motion, inertial frame 86
Free fall from tower 86
Uniform linear motion, rotating frame 87
Centrifugal correction to g 88
Direction of winds 88

Mathematical Note 1: Differentiation of products of vectors 89

Mathematical Note 2: Angular velocity as a vector quantity 89

Historical Note: Rotating bucket—Newton's view 90

Galilean Invariance

Review. Newton's Laws of Motion. This chapter is chiefly concerned with Newton's laws of motion. You have considered these laws in high school. We review them first for bodies of constant mass.

Newton's first law. A body remains in a state of rest or constant velocity (zero acceleration) when left to itself, with no external force acting on it. That is,

$$\mathbf{a} = 0 \qquad \text{when} \qquad \mathbf{F} = 0.$$

Newton's second law. The total force on a body is the product of the mass of the body times the acceleration:

$$\mathbf{F} = M\mathbf{a}.$$

Newton's third law. Whenever two bodies interact, the force $\mathbf{F}_{12}$ on the second body (2) due to the first (1) is equal and opposite to the force $\mathbf{F}_{21}$ on the first (1) due to the second (2):

$$\mathbf{F}_{12} = -\mathbf{F}_{21}.$$

There are inherent limitations to the validity of the third law: we believe, as we shall discuss in Chap. 10, that all signals or forces have a finite propagation velocity. The third law, however, states that $\mathbf{F}_{12}$ is equal and opposite to $\mathbf{F}_{21}$ when both are measured *at the same time.* This requirement is inconsistent with the finite time interval required for one particle to feel the force the second particle is exerting. In atomic collisions the third law is therefore not always a good approximation. In automobile collisions it is quite a good approximation, because the duration of the collision is long in comparison with the time it takes a light signal to traverse a crumpled automobile. This time is of the order of

$$\frac{L}{c} \approx \frac{300 \text{ cm}}{3 \times 10^{10} \text{ cm/sec}} \approx 10^{-8} \text{ sec},$$

where L is the length of the crumpled automobile. (In 10^{-8} sec a car moving at 100 km/hr, which is about 3×10^3 cm/sec, travels about 3×10^{-5} cm.)

The first two laws hold only when observed in unaccelerated reference frames. This is obvious from everyday experience. If your reference frame is at rest on a rotating merry-go-round, you do not have zero acceleration in this frame in the absence of applied forces. You can only stand still on the merry-go-round by pushing out to cause on your body a force $M\omega^2 r$ toward the axis, where M is your mass, ω the angular velocity, and r your distance from the axis of rotation. Or suppose your reference frame is at rest in an aircraft which accelerates rapidly on takeoff. You are pressed back against your seat by the acceleration, and you are held at rest by the force exerted on you by the back of the seat.

If you were content to remain in uniform motion or at rest relative to an unaccelerated reference frame, no force would be needed. But if you insist on being at rest in an accelerated reference frame, you must exert or experience forces—you need a rope to hold onto or a seat to back you up. The forces which arise automatically in accelerated reference frames are important in physics. It is particularly important to understand the forces that act in a reference frame which is in circular motion. It is a good idea to review now this topic which you studied in high school. There is an excellent PSSC film (PSSC-MLA 0307) dealing with inertial reference frames. The film clarifies beautifully the central material of this chapter.

Example. Ultracentrifuge. The effects of *not* being in an inertial reference frame can be enormous! And the effects can be of great practical importance. Consider a molecule suspended in a liquid in the test chamber of an ultracentrifuge. Suppose that the molecule lies 10 cm from the axis of rotation and that the ultracentrifuge rotates at 1000 revolutions per second (60,000 rpm). Then the angular velocity is

$$\omega = (2\pi)(1 \times 10^3) \cong 6 \times 10^3 \text{ radians/sec},$$

and the linear velocity is

$$v = \omega r \cong (6 \times 10^3)(10) \cong 6 \times 10^4 \text{ cm/sec.}$$

The magnitude of the acceleration associated with circular motion is equal to $\omega^2 r$ (we saw this in Chap. 2):

$$a = \omega^2 r \cong (6 \times 10^3)^2 (10) \cong 4 \times 10^8 \text{ cm/sec}^2.$$

Now the acceleration g due to gravity is only 980 cm/sec² at the surface of the earth, so that the ratio of the rotational acceleration to the gravitational acceleration is

$$\frac{a}{g} \cong \frac{4 \times 10^8}{10^3} \cong 4 \times 10^5.$$

Thus the acceleration in the ultracentrifuge is 400,000 times as large as the acceleration due to gravity. (These data are characteristic of the ultracentrifuge shown in the photograph.) Suspended molecules *whose density (mass/volume) is different* from that of the surrounding liquid will experience in the ultracentrifuge cell a strong centrifugal force tending to separate them from the fluid. If their density is the same as the liquid, there is no separation effect.

According to Newton's first law the suspended molecule wants to remain at rest (or moving at constant velocity in a straight line) as viewed from the laboratory. (The laboratory is a fairly good approximation to an unaccelerated reference frame.) The molecule does not want to be dragged around madly at a high angular velocity in the ultracentrifuge. To an observer at rest in the ultracentrifuge the molecule will act as if there were exerted on it a force $M\omega^2 r$ tending to pull it away from the axis of rotation, toward the outside of the test chamber in the centrifuge rotor. How big is the force? Suppose the molecular weight of the molecule is 100,000, which means roughly that the mass M of the molecule is 10^5 times the mass of a proton:

$$M \cong (10^5)(1.7 \times 10^{-24}) \cong 2 \times 10^{-19} \text{ gm.}$$

(The mass of a proton is approximately equal to one atomic mass unit, as we see from the table of values inside the cover.) The force associated with the rotational acceleration is

$$Ma = M\omega^2 r \cong (2 \times 10^{-19})(4 \times 10^8) \cong 8 \times 10^{-11} \text{ dynes,}$$

Rotor of an ultracentrifuge. This operates at 60,000 rpm and provides a centrifugal acceleration slightly below 300,000 times the acceleration of gravity. (*Beckmann Spinco Division*) using the value of the acceleration given in the preceding paragraph. This force, which tends to pull the molecule toward the outside of the test chamber, is called the *centrifugal* (center-fleeing) force. The motion toward the outside is opposed by the drag of the surrounding liquid on the molecule. Since different kinds of molecules will experience different values of the centrifugal force and different drags, they will accordingly move outward in the test cell at different speeds. The ultracentrifuge thus provides an excellent method for the separation from one another of different kinds of molecules. The method works best on large molecules, which tend to be just the molecules of great biological interest, and so the matter of whether a molecule is at rest in an accelerated, or an unaccelerated, reference system turns out to be important for biological and medical research.

Inertial Reference Frames. The fundamental law of classical mechanics is Newton's second law of motion:

$$\text{Force} = \frac{d}{dt}(\text{momentum}); \qquad \mathbf{F} = \frac{d}{dt}\mathbf{p}, \tag{1}$$

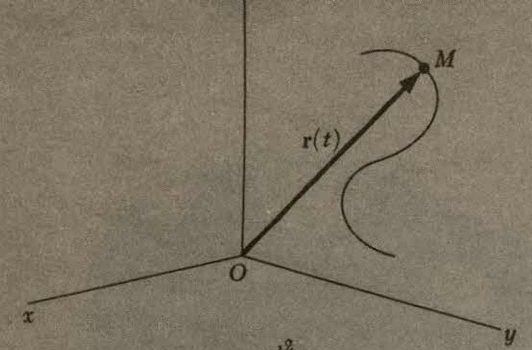
where $\mathbf{F}$ is the force and $\mathbf{p} \equiv M\mathbf{v}$ is the momentum. This law is valid for an observer stationed in an unaccelerated reference frame. Such a reference frame is called an *inertial reference frame*, or inertial reference system. For a body of constant mass M we have

$$\mathbf{F} = M\frac{d\mathbf{v}}{dt} \equiv M\frac{d^2\mathbf{r}}{dt^2} \equiv M\mathbf{a}, \tag{2}$$

where $\mathbf{a}$ is the acceleration. Sometimes the mass is not constant; it is not constant for a satellite that is ejecting fuel nor for any body in the region of relativistic velocities (close to the velocity of light). Equation (1) or (2) may be viewed as defining in a consistent way the *true force* $\mathbf{F}$ which acts on a system.

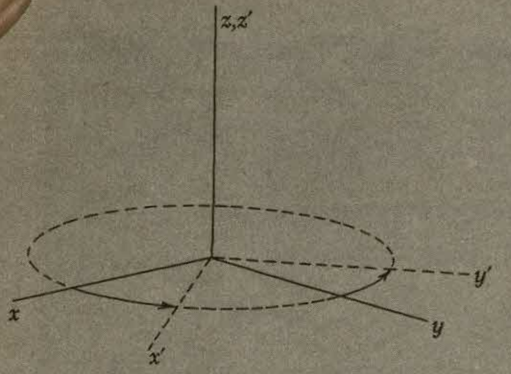
Does a laboratory fixed on the surface of the earth provide a good inertial reference frame? If it does not, how do we correct $\mathbf{F} = M\mathbf{a}$ to allow for the acceleration of the laboratory?

For many purposes the earth is fairly good approximation to an inertial frame. An acceleration of a laboratory fixed on the earth results from the daily rotation of the earth about its axis. This rotation amounts to a small acceleration of the laboratory, not entirely negligible for all purposes. A point at rest on the surface

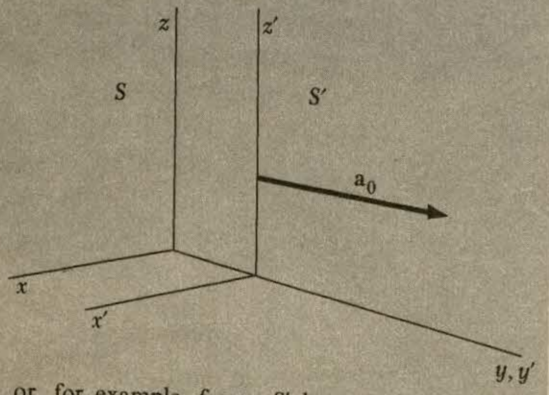


$$\mathbf{F} = M\frac{d^2\mathbf{r}}{dt^2}$$

Newton's Second Law says:
Force = mass × acceleration.
But, acceleration relative to what frame?



For instance frame $S'(x',y',z')$ rotates with respect to frame $S(x,y,z)$. The acceleration of M in each of these frames is different . . .



or, for example, frame S' has acceleration $\mathbf{a}_0$ with respect to frame S. The acceleration of M in each of these frames is different.

of the earth at the equator must experience a centripetal acceleration given by

$$a = \frac{v^2}{R_E} = \omega^2 R_E, \tag{3}$$

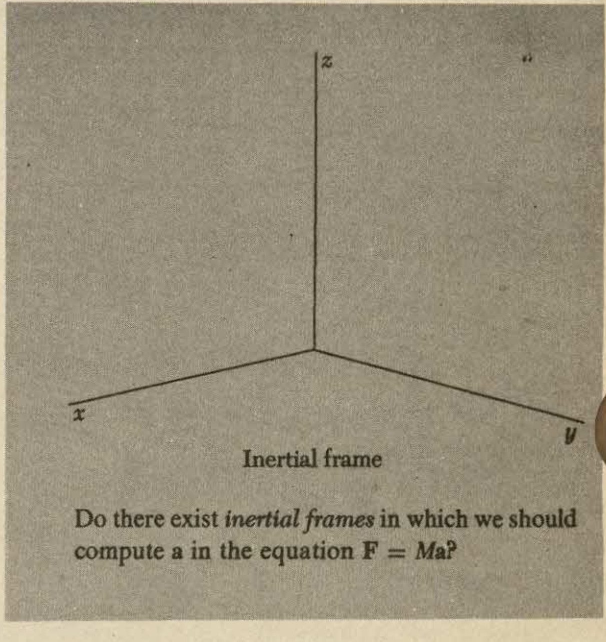
referred to the center of the earth. Hence $\omega = 2\pi f$ is the angular velocity of the earth and R_E is the radius of the earth. The earth turns through 2π radians in 1 day. (There are about 8.6×10^4 sec in one day.) The angular velocity of the earth is therefore

$$\omega = \frac{2\pi}{8.6 \times 10^4} \approx 0.73 \times 10^{-4} \text{ sec}^{-1}. \tag{4}$$

With $R_E \cong 6.4 \times 10^8$ cm, the acceleration is

$$a \approx (0.73 \times 10^{-4})^2 (6.4 \times 10^8) \approx 3.4 \text{ cm/sec}^2. \tag{5}$$

This is a large part of the amount by which the observed acceleration of gravity ($g \approx 980$ cm/sec^2) at the North Pole exceeds that at the Equator. The remainder of the large-scale gravity variation on the surface of the earth is due to the ellipsoidal shape of the earth. The total variation between the North Pole (or South Pole) and the Equator is about 5.2 cm/sec^2. Until the availability of satellites the best way to determine the flattening of the earth at the poles was by measuring the variation of gravity over the earth.



Inertial frame

Do there exist *inertial frames* in which we should compute a in the equation $\mathbf{F} = M\mathbf{a}$?

Values of g at diverse latitudes

Station	*Latitude*	*g, cm/sec^2*
North Pole	90°N	983.245
Karajak Glacier, Greenland	70°N	982.53
Reykjavik, Iceland	64°N	982.27
Leningrad	60°N	981.93
Paris	49°N	980.94
New York	41°N	980.27
San Francisco	38°N	979.96
Honolulu	21°N	978.95
Monrovia, Liberia	6°N	978.16
Batavia, Java	6°S	978.18
Melbourne, Australia	38°S	979.99

Later in this chapter we shall construct a more complicated form of Newton's second law which is applicable in a coordinate system with axes fixed on the surface of the earth. But to get a valid law in the simple form (1) or (2) *we must refer the acceleration to a reference system which is unaccelerated.* Such a reference system is called *inertial*, or Galilean. In an accelerated (noninertial) reference system **F** does not equal $M\mathbf{a}$ if **a** is the acceleration as observed from the noninertial system.

It is an established convention to speak of the fixed stars as a standard unaccelerated reference system. This language contains an element of metaphysics, for the statement that the fixed stars are unaccelerated goes beyond our actual experimental knowledge. It is unlikely that our instruments could detect an acceleration of a distant star or group of stars of less than 10^{-4} cm/sec^2, even if we made careful observations for 100 years. For practical purposes it is convenient to refer directions in space to the stars, but also for practical purposes we shall see that we can establish by experiment a satisfactory unaccelerated reference system. Even if the earth were surrounded continuously by a dense fog, we would be able to establish an inertial reference frame without particular difficulty.

The acceleration of the earth in its orbit around the sun is one order of magnitude smaller than the acceleration due to the rotation of the earth. Since 1 year $\cong 3 \times 10^7$ sec, the angular frequency of the earth about the sun is

$$\omega \approx \frac{2\pi}{3 \times 10^7} \approx 2 \times 10^{-7}\ \text{sec}^{-1}. \tag{6}$$

We have, with $R \cong 1.5 \times 10^{13}$ cm, the centripetal acceleration

$$a = \omega^2 R \approx (4 \times 10^{-14})(1.5 \times 10^{13}) \approx 0.6\ \text{cm/sec}^2, \tag{7}$$

for the acceleration of the earth in its orbit about the sun.

The acceleration of the sun toward the center of our galaxy† is not known experimentally. But from Doppler-shift studies of spectral lines (Chap. 10) the velocity of the sun relative to the center of the galaxy is believed to be about 3×10^7 cm/sec. If the sun is in a circular orbit about the center of the galaxy, which is at a distance of approximately 3×10^{22} cm from the sun, then the acceleration of the sun about the galactic center is

---

†Stars are not scattered randomly throughout space but are gathered into large systems widely separated from each other. Each system contains of the order of 10^{10} stars. The systems are called galaxies; that one which contains our own sun is known as the Galaxy. The Milky Way is part of our galaxy. The galaxies themselves are not distributed entirely at random, for there is a marked tendency to form clusters. Our galaxy belongs to a cluster of 19 members known as the Local Group which forms a bound physical system, bound by gravitational attraction.

$$a = \omega^2 R = \frac{v^2}{R} \approx \frac{9 \times 10^{14}}{3 \times 10^{22}} \approx 3 \times 10^{-8} \text{ cm/sec}^2. \qquad (8)$$

This is quite small. We do not know by observation that the sun is not accelerating much faster than this, nor do we know that the center of the galaxy is not itself accelerating significantly.

We do know in practice that the set of assumptions central to classical mechanics works out exceedingly well. These assumptions are:

1. Space is Euclidean.
2. Space is isotropic, so that physical properties are the same in all directions in space. Thus the mass M in $\mathbf{F} = M\mathbf{a}$ does not depend on the direction of $\mathbf{a}$.
3. Newton's laws of motion hold in an inertial system determined for an observer at rest on the earth by taking account only of the acceleration of the earth about its axis and in its orbit about the sun.
4. Newton's law of universal gravitation is valid. This law states that there exists an attractive force $F = GM_1M_2/R^2$ between any two point masses M_1, M_2 at separation R; here G is a constant.

These assumptions are difficult to test individually to great precision. The most precise tests, which relate to the motions of the planets in the solar system, usually involve the entire package of all four statements above. Several extremely precise tests of the classical package are discussed in the historical notes at the end of Chap. 5.

Forces in Inertial Reference Systems. Galileo said that *a body subject to no forces has a constant velocity.*† We have seen that this statement is true only in an inertial reference system—it defines an inertial system.

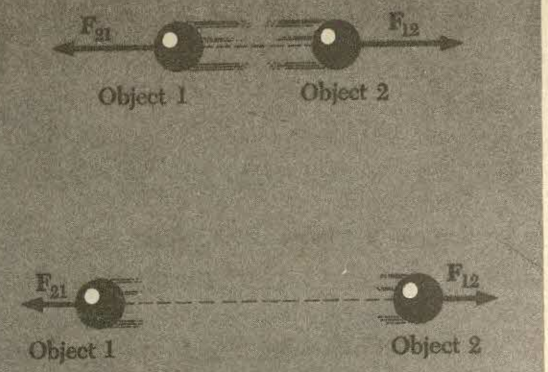
The statement may seem ambiguous, for how do we ever know that a body has no forces acting upon it? Forces can act upon a body not only by the direct contact of one body with another, but forces can act also when the body is isolated. Gravitational or electrical forces can be important even without the very near presence of other bodies. We cannot be sure that no forces act just because no other bodies touch or are very close to a given body. But if we cannot decide *a priori* that some reference body is not subject to a force, we have difficulty framing laws of motion which relate forces to accelerations. We need to have an unaccelerated reference system with respect to which we can measure accelerations; Galileo's way of determining such a system assumes that we have some independent way of knowing that there is no force upon it. But we don't know this, because our criterion for no force is no acceleration, which demands some reference against which to measure acceleration, and so on around the circle of argument.

---

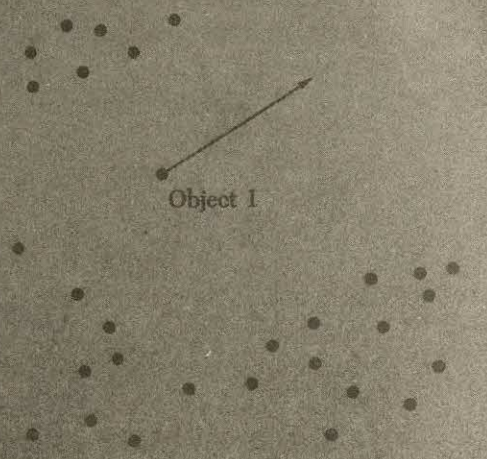
† This is often called Newton's first law of motion.

The situation is not hopeless, for we know the forces between two bodies must fall off fairly rapidly as the distance between the bodies grows. If the forces did not fall off rapidly, we could never isolate the interactions of two bodies from among those of all other bodies in the universe. All known forces between particles fall off with distance at least as fast as that of the inverse-square law. We and every other body on the earth are attracted mainly toward the center of the earth and not mainly toward some distant part of the universe. If we were not supported by the floor, we would accelerate toward the center at 980 cm/sec^2. We are pulled less strongly by the sun; from Eq. (7), we accelerate toward it at 0.6 cm/sec^2. In a reasonable description of acceleration, a body far away from all other bodies is expected to have virtually no forces exerted upon it and hence have no acceleration. A typical star is at least 10^{18} cm from its nearest neighbor† and is expected to have only a small acceleration. We are thus led to expect that the fixed stars may define a convenient unaccelerated coordinate system, to a good approximation.

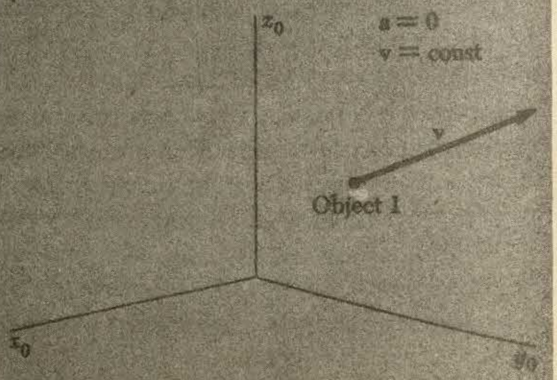
A good discussion of the establishment of an unaccelerated reference system is given by P. W. Bridgman, *Am. J. Phys.* **29**, 32 (1961). Several excerpts follow: "A system of three rigid orthogonal axes fixes a Galilean frame if three force-free massive particles projected along the three axes with arbitrary velocities continue to move along the axes with uniform velocities. Our terrestrial laboratories do not constitute such a frame, but we may construct such a frame in our laboratories by measuring how three arbitrarily projected masses deviate from the requirement . . . and incorporating these deviations as negative corrections into our specifications for the Galilean frame. There need be no reference to the stars, but the behavior of bodies can be relevantly described in terms of such immediately observable things as the rotation of the plane of the Foucault pendulum with respect to the earth or the deviation of a falling body from the perpendicular. Even if



Experimentally, the force one object exerts on another always decreases rapidly as the objects are separated to greater distance. Thus . . .



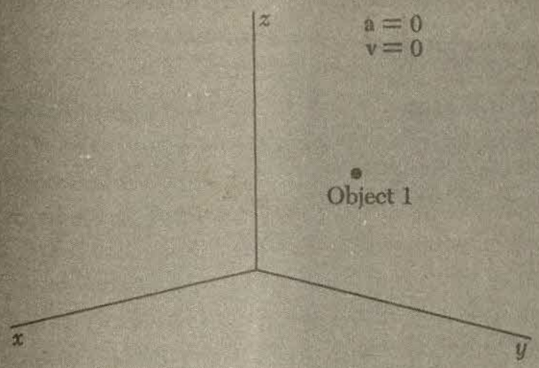
if object 1 is sufficiently far from all other objects, it will be subjected to no forces.



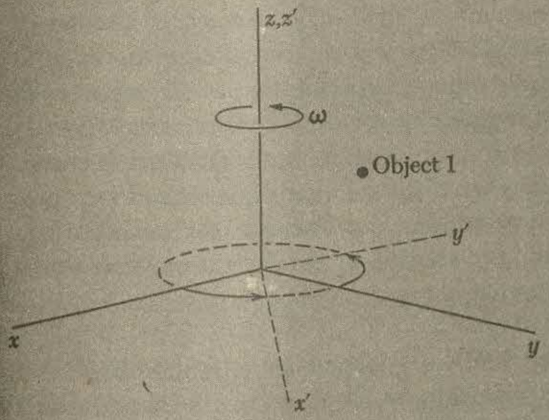
An *inertial frame* is one in which the acceleration of an object like 1 is zero.

---

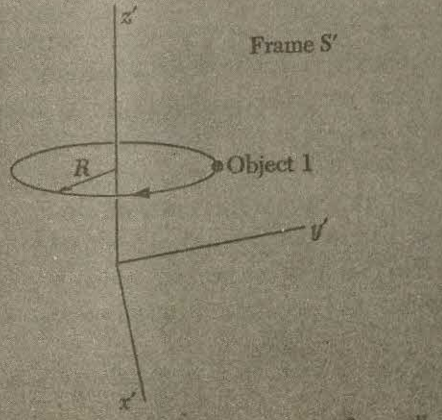
† Double (binary) stars are another matter.



In particular there are inertial frames in which object 1 is, and remains, at rest.



If $S(x,y,z)$ is such an inertial frame, then $S'(x',y',z')$, which rotates around the z axis of S, *cannot* be inertial. For . . .



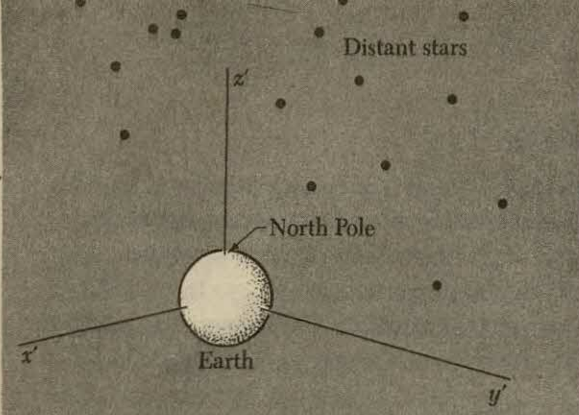
In frame S' object 1, although very far from all other objects, experiences acceleration. (It appears to rotate.) For example . . .

the rocket operator who is trying to put a satellite into orbit finds it convenient to make some of his specifications in terms of observations on the pole star, it is obvious that his apparatus must eventually be described in terrestrial terms. . . . In a Galilean frame a rotating body, after it has been set into rotation and the forces disconnected, preserves the orientation of its plane of rotation in the frame and, consequently, preserves the direction of its axis of rotation."

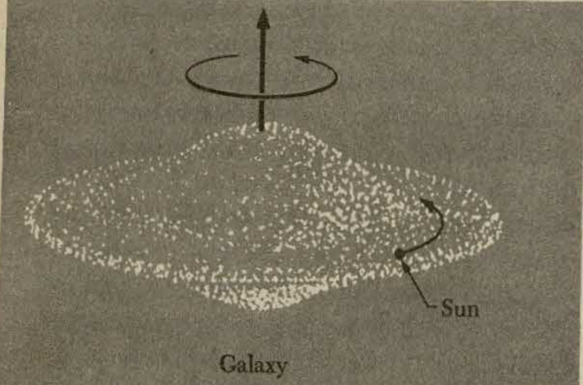
Absolute and Relative Acceleration. An inertial reference system can be found in which **F** equals $M\mathbf{a}$ to very great accuracy. This is well-supported by experiment. We conclude that in an inertial reference system the forces which have been postulated to explain the motion of galaxies, stars, atoms, electrons, etc., have the common property that the force on a body does indeed decrease as it is removed further and further from its neighbors. We shall see that if we choose a noninertial reference system there *appear* to exist forces which do not have the property of being associated with the proximity of other bodies.

The existence of an inertial reference system suggests a difficult and unanswered question: What effect does all of the other matter in the universe have upon an experiment done in a terrestrial laboratory? Suppose, for example, all the matter in the universe except that in the neighborhood of our own earth were to be given a large acceleration **a**. A particle on earth, subject to no net force originally, had zero acceleration relative to the fixed stars. When these stars are accelerated, would this particle—originally force-free—still maintain zero acceleration with respect to the nonaccelerated stars, or would it experience a change in its motion relative to its immediate surroundings? Is there any difference between accelerating a particle with $+\mathbf{a}$ or the fixed stars with $-\mathbf{a}$? If only the relative acceleration is significant, the answer to the last question is *no;* if there is a meaning to absolute acceleration, the answer is *yes*. This is a fundamental unanswered question, but it is not easily susceptible to experimental investigation.

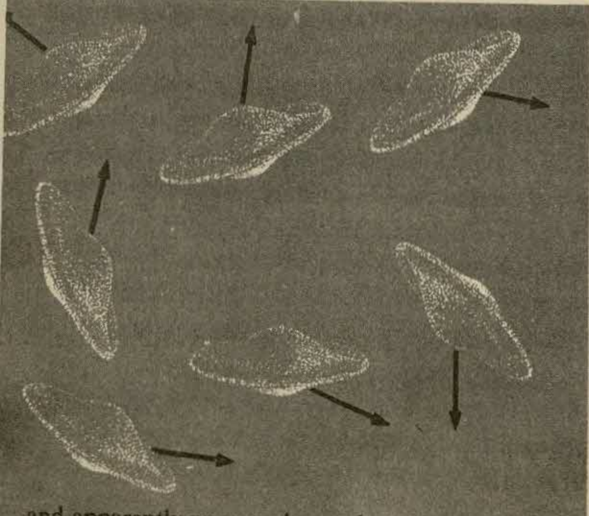
Newton expressed this question and his own answer in a picturesque way. Consider a bucket of water. If we rotate the bucket relative to the stars, the water surface assumes a parabolic shape; on this everyone would agree. But suppose that instead of rotating the bucket we somehow rotated the stars about the bucket so that the relative motion was the same. Newton's belief was that the surface would be flat if we rotated the stars. This



from frame $S'(x',y',z')$ fixed on earth, the distant stars, which are like object 1, rotate. A frame fixed on earth is not inertial because the earth spins on its axis and revolves about the sun.



Is a frame fixed on the sun inertial? Even the sun rotates around the galaxy, but this acceleration appears to be small enough to neglect . . .



and apparently we can also neglect the acceleration of our galaxy relative to others.

viewpoint gives a significance to absolute rotation and absolute acceleration. What we do know empirically is that all the phenomena of the rotating bucket of water can be completely described and correlated with the results of local measurements in the laboratory, with no reference whatever to the stars.

The opposite point of view, that *only acceleration relative to the fixed stars has any significance,* is a conjecture commonly called Mach's principle. Although there is neither experimental confirmation nor objection to this point of view, some physicists including Einstein have found this principle to be attractive *a priori.* Others have not found it attractive. This is a matter for speculative cosmology.

If one believes that the average motion of the rest of the universe affects the behavior of any single particle, a number of related questions present themselves without offering any clues to the answers. Are there other relations between the properties of a single particle and the state of the rest of the universe? Would the charge on the electron, or its mass, or the interaction energy between nucleons† change if the number of particles in the universe or their density were somehow altered? So far the answer to this deep question, the relation of the distant universe to the properties of single particles, remains unanswered.

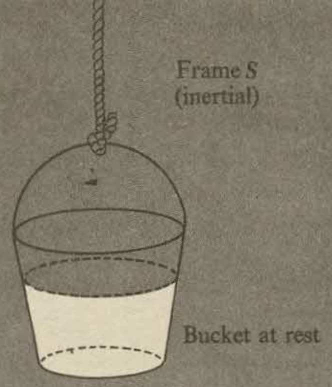
Absolute and Relative Velocity. Is there any physical meaning to absolute velocity? According to all experiments yet performed the answer is *no.* We are thus led to a fundamental hypothesis, the hypothesis of *Galilean invariance:*

> *The basic laws of physics are identical in all reference systems which move with uniform (unaccelerated) velocity with respect to one another.*

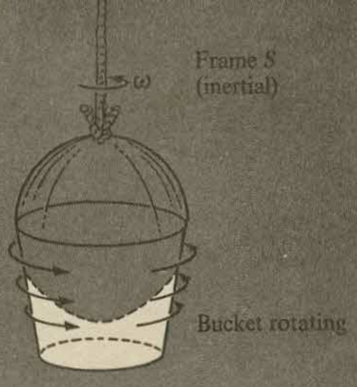
According to this hypothesis an observer confined to a windowless box cannot tell by any experiment whether he is stationary or in uniform motion with respect to the fixed stars. Only by looking through a window, so that he can compare his motion to that of the stars, can an observer tell that he is in uniform motion with respect to them. Even then he could not decide whether he or the stars were moving. The Galilean invariance principle was one of the first to be introduced into physics. It was basic to Newton's

---

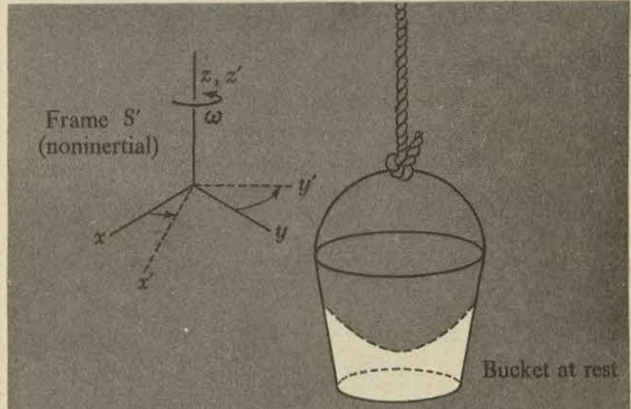
† A nucleon is a proton or a neutron; an antinucleon is an antiproton or antineutron.



An example of "fictitious" forces which arise in noninertial frames: when the bucket is at rest in S, the water surface is flat. S is assumed to be unaccelerated relative to distant stars.



When the bucket rotates in S, the water surface assumes a paraboloidal shape.

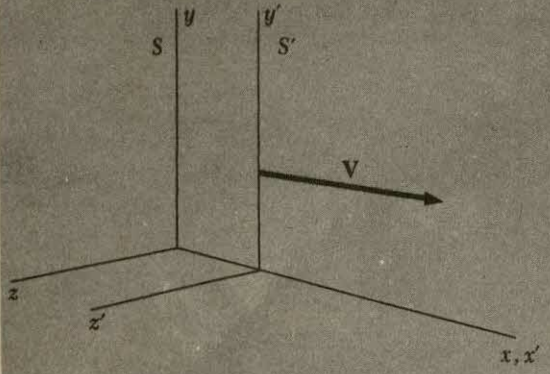


In rotating frame S', the bucket is at rest. But the water surface is still paraboloidal! A "fictitious" centrifugal force acts on the water in the noninertial frame S'.

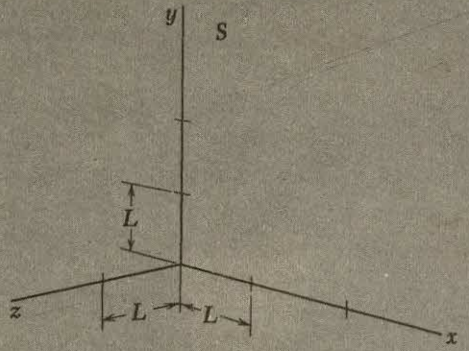
view of the universe; it has survived repeated experiments, and it is one of the cornerstones of the theory of special relativity. It is such a remarkably simple hypothesis that it would be considered seriously even in the absence of strong evidence. The hypothesis of Galilean invariance is entirely consistent with special relativity, as we shall see in Chap. 11.

What use can we make of the hypothesis? The hypothesis that absolute velocity has no meaning in physics restricts in part the form and content of all physical laws, both known and undiscovered. To two observers moving with different velocities, but without relative acceleration, the laws of physics must be the same if the hypothesis is true. Suppose they both observe some particular phenomenon, such as the collision of two particles. Because of the different velocities of the observers, the observed event will be described differently by each of them. From the laws of physics we can predict what the observations of one observer will be, how the particles interact, and finally, how they appear to the second observer.

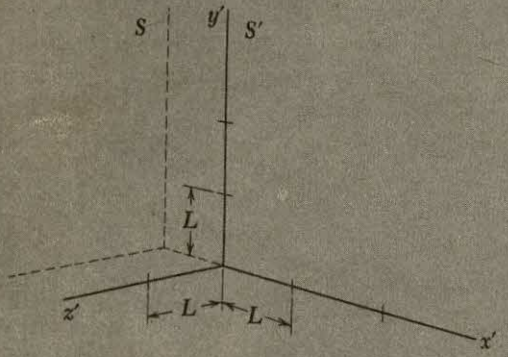
The laws of physics of the second observer can therefore be obtained from those of the first by two separate lines of argument. By hypothesis they are the same as those of the first. Alternately, we can predict the second observer's laws from what we predict about his observations of the phenomena described by the first observer's laws. The two methods give the same result for actual physical laws. Before proceeding, we state some empirical results on the manner in which two observers, one moving at uniform velocity with respect to the other, describe the same physical event.



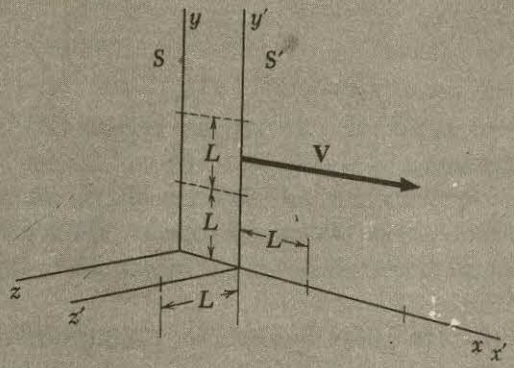
Suppose S is an inertial frame and S′ moves with constant velocity **V** relative to S. Then S′ must be inertial also.



Let us mark off equal lengths L along the (xyz) axes of S . . .



and along the (x',y',z') axes of S′.



Then to an observer in S, the lengths in S′ appear unaltered even though S is moving according to the Galilean transformation.

Galilean Transformation. If we now discuss how two observers measure a given length and time interval, we can infer how their measurements will compare for other physical quantities. Let S denote a particular inertial Cartesian coordinate system, and let S′ denote another inertial Cartesian coordinate system moving with velocity **V** with respect to the first. The axes x', y', z' of S′ are taken to be parallel to the x, y, z axes of S. We choose **V** to be in the x direction. We wish to compare measurements of time and distance by an observer sitting in the frame S′ with those of an observer at rest in the frame S. The result of the comparison ultimately can only be decided by experiment.

If our two observers construct identical clocks, they may perform the following experiment: We suppose first that the observer on S distributes his clocks along his x axis and sets them all to read the same. This is not necessarily a simple operation to carry out, and we postpone analysis of exactly how such measurements are to be made until we discuss in Chap. 11 the analogous experiment in the theory of special relativity. If, however, we approximate the speed of light as infinite,[†] we need only look at all the clocks to be sure they have the same initial setting. Now we can compare the reading of clocks S′ with clocks 1, 2, 3, . . . in S as S′ passes by each of them. If such an experiment is to be done with a real macroscopic clock, we are for practical technical reasons restricted to a velocity V for S′ of the order of 10^6 cm/sec, which is a typical satellite velocity. In this regime $V/c \ll 1$ and experiment confirms that if the clock on S′ is set to agree with clock 1, it will agree with 2, 3, 4. . . . As accurately as we can measure[‡] in these conditions we assert that

$$t' = t. \tag{9}$$

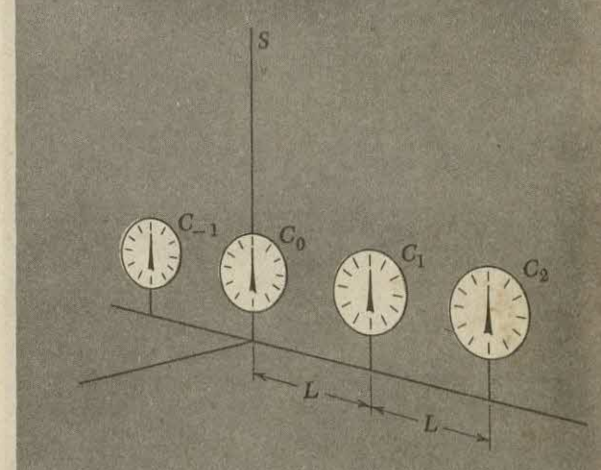
That is, times read in S′ are equal to times read in S. Here t refers to the time of an event in S and t' to an event in S′.

This result is neither self-evident nor is it exactly true for all velocities V, as we shall see in Chap. 11. We can also determine the relative sizes of a stationary and a moving meter stick. We want to know the apparent size to the observer on S of a meter stick at rest in S′. A simple way to find this is to utilize the clocks again and record the positions of both ends of the moving meter stick simultaneously, that is, when the clocks on S at the front and back read the same. We find by experiment[§] that

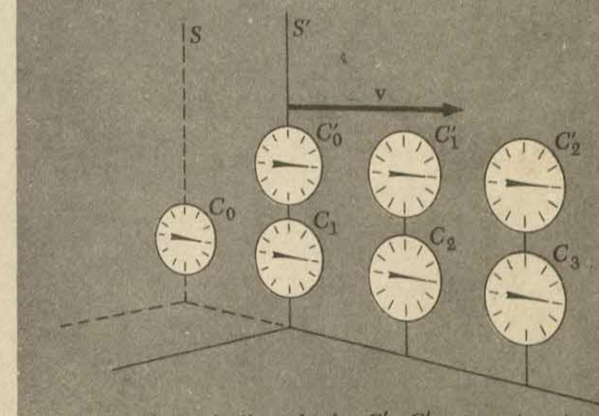
$$L' = L, \tag{10}$$

provided that $V \ll c$.

We can summarize (9) and (10) in terms of a transformation which relates the coordinates x', y', z' and time t' as measured on S′ to the coordinates x, y, z and t as measured on S. The frame S′



Let us place synchronized clocks C_0, C_1, etc., at intervals of length L along x axis, at rest in S.



If we place similar clocks C'_0, C'_1, etc., at rest in S′, then to an observer in S these clocks appear synchronized with themselves *and* with C_0, C'_1, etc., according to the Galilean transformation.

[†] This procedure can be simply improved upon by correcting for the time it takes for the image of the more distant images to reach our eyes, so that a clock which is l cm away will appear to lag behind a nearby clock by l/c sec, where $c \cong 3 \times 10^{10}$ cm/sec is the velocity of light.

[‡] Relativity theory predicts for a velocity $V = 10^6$ cm/sec that t' and t should differ by only one part in 2×10^9, or less than 1 sec in 50 years. Although clocks with such stability can now be constructed, there was no way until the launching of satellites to keep a clock moving at 10^6 cm/sec for enough time to permit a measurement. The equality $t = t'$ for $V \ll c = 3 \times 10^{10}$ cm/sec is a simple extrapolation from experience and is not based to date upon very accurate measurement.

[§] Such an experiment has not been performed with great accuracy, and the belief in the equality $L = L'$ for $V \ll c$ is based mainly upon qualitative experience, the simplicity of the hypothesis, and the fact that this assumption does not lead to any paradoxes or inconsistencies.

moves with velocity $V\hat{\mathbf{x}}$ as viewed from S. Suppose at $t = 0$ that $t' = 0$, and that at this time the origins O and O' coincide. If we choose identical scales for distance, we will have the following transformation equations:

$$\boxed{t = t'; \qquad x = x' + Vt'; \qquad y = y'; \qquad z = z'.} \tag{11}$$

This transformation is called a Galilean transformation.

An immediate consequence of (11) is the law of velocity addition:

$$v_x = \frac{x}{t} = \frac{x}{t'} = \frac{x'}{t'} + V = v_x' + V, \tag{12}$$

or, in vector form,

$$\mathbf{v} = \mathbf{v}' + \mathbf{V}, \tag{13}$$

where $\mathbf{v}'$ is the velocity measured in S′, and $\mathbf{v}$ is the velocity measured in S. The inverse transformation to (13) is simply $\mathbf{v}' = \mathbf{v} - \mathbf{V}$.

If the definition (11) of a Galilean transformation between S and S′ is combined with the fundamental postulate that the laws of physics are identical as determined by physicists on S and S′, then we can make the following statement:

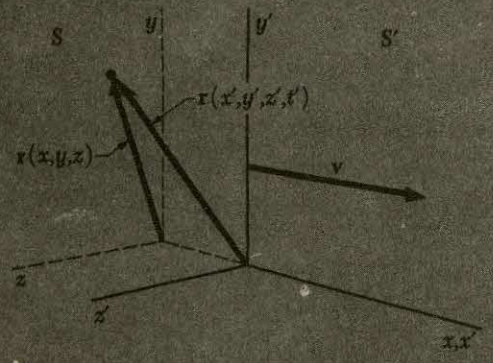
> The basic laws of physics are unchanged in form in two reference frames connected by a Galilean transformation.

This statement is somewhat more special than our earlier general statement that the laws of physics are identical in all reference frames which move with uniform velocity with respect to one another. The more special statement involves, for example, the assumption that $t = t'$. We shall see in Chap. 11 that this assumption will have to be modified. The general principle of invariance is believed always to be valid, but the appropriate exact transformation equations are the Lorentz transformation equations and not the equations (11).

The present assumption of invariance under (11) means that the laws of physics must have exactly the same appearance when written in the primed and in the unprimed variables, as in (16) to (18) below. This requirement puts a definite restriction on the possible form of physical laws.

From the relation $\mathbf{v} = \mathbf{v}' + \mathbf{V}$, where $\mathbf{V}$ is the relative velocity of the two reference frames, it follows that

$$\Delta\mathbf{v} = \Delta\mathbf{v}'; \tag{14}$$



$$\mathbf{r}(x',y',z',t') = \mathbf{r}'(x - vt, y, z, t)$$

We can thus summarize the Galilean transformation from S ↔ S′:

$$x' = x - Vt \qquad z' = z$$
$$y' = y \qquad t' = t.$$

a velocity change observed from S is equal to a velocity change observed from S'. Both S and S' are inertial frames. We recall that $\mathbf{V}$ is assumed not to change with time. Because $\Delta t = \Delta t'$ it follows that the accelerations are equal as observed from S and S':

$$\mathbf{a} \equiv \frac{\Delta \mathbf{v}}{\Delta t} = \frac{\Delta \mathbf{v}'}{\Delta t'} \equiv \mathbf{a}'. \tag{15}$$

How does the force $\mathbf{F}$ transform from S to S'? The assumption that the laws of physics are the same in the primed as in the unprimed variables means that

$$\mathbf{F}' = M\mathbf{a}' \tag{16}$$

if

$$\mathbf{F} = M\mathbf{a}, \tag{17}$$

provided that the mass M is independent of velocity. But we have shown in (15) that $\mathbf{a}' = \mathbf{a}$, whence

$$\mathbf{F} = M\mathbf{a}' = \mathbf{F}', \tag{18}$$

and thus the forces are equal: $\mathbf{F} = \mathbf{F}'$. We conclude that if the relation $\mathbf{F} = M\mathbf{a}$ is used to define the force, observers in all inertial reference frames would agree on the magnitude and direction of the force $\mathbf{F}$ independent of the relative velocities of the reference frames.

Conservation of Momentum. Among the physical laws which are compatible with Galilean invariance, the conservation of momentum and the conservation of mass and energy have a particular importance. These laws are already familiar in your earlier physics, where they were introduced without any reference to invariance. The law of conservation of energy states that the total energy of the universe is constant, independent of time. The law of conservation of mass states that the total mass of the universe is constant, independent of time. We shall not discover anything we did not already know by considering these laws from the invariance point of view. But we shall have gained an insight which will help us to generalize the law of momentum conservation to relativistic situations where $\mathbf{F} = M\mathbf{a}$ is no longer an exact law of nature. Our ultimate problem will be to discover the equivalents of mass, energy, and momentum conservation in the regime of relativistic speeds.

The law of conservation of momentum as applied to the collision of two bodies, 1 and 2, states that the sum of the momenta after the collision is equal to the sum of the momenta before the collision:

$$\mathbf{p}_1(\text{before}) + \mathbf{p}_2(\text{before}) = \mathbf{p}_1(\text{after}) + \mathbf{p}_2(\text{after}), \qquad (19)$$

where the momentum $\mathbf{p}$ is defined by

$$\mathbf{p} \equiv M\mathbf{v}. \qquad (20)$$

We assume that the collision takes place in a region of space away from external forces. The collision may be either elastic or inelastic. In an elastic collision all the kinetic energy of the incoming particles reappears after the collision as the kinetic energy of the same particles. In the usual inelastic collision, part of the kinetic energy of the incoming particles appears after the collision as some form of *internal* excitation energy (such as heat) of one or more of the particles. It is important to realize that we can apply momentum conservation *even* to an *inelastic* collision.

The law of conservation of mass as applied to the collision of two bodies states that the sum of the masses after the collision is equal to the sum of the masses before the collision:

$$(M_1 + M_2)_{\text{before}} = (M_1 + M_2)_{\text{after}}. \qquad (21)$$

In Chap. 12 both (20) and (21) will be generalized to relativistic velocities.

We assume for the present that the mass of each particle is conserved in a collision. We now give two separate derivations of the law of conservation of linear momentum. The first derivation is based on the assumption of *Newtonian forces.* The second derivation is better and more general; it is based on the assumptions of *Galilean invariance* and *conservation of energy.*

First derivation. The conservation of momentum is not a necessary consequence of $\mathbf{F} = M\mathbf{a}$ alone. The present derivation assumes that the forces among particles are of a special sort called *Newtonian* which satisfy Newton's third law. A Newtonian force is defined by the property that the force $\mathbf{F}_{12}$ which one particle (1) exerts upon the other (2) is the negative of the force $\mathbf{F}_{21}$ exerted by 2 on 1:

$$\mathbf{F}_{12} = -\mathbf{F}_{21}. \qquad (22)$$

From Newton's second law we have during any infinitesimal interval of time Δt that

$$\mathbf{F}_{12} = M_2\frac{\Delta\mathbf{v}_2}{\Delta t}; \qquad \mathbf{F}_{21} = M_1\frac{\Delta\mathbf{v}_1}{\Delta t}. \qquad (23)$$

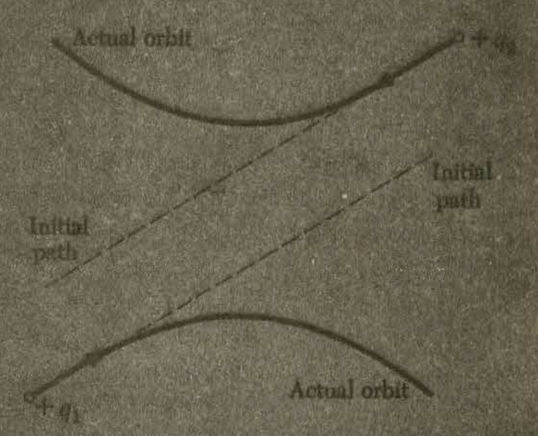
But $\mathbf{F}_{12} = -\mathbf{F}_{21}$, so that we have the law of conservation of momentum

$$M_1\,\Delta\mathbf{v}_1 + M_2\,\Delta\mathbf{v}_2 = 0, \tag{24}$$

or

$$(M_1\mathbf{v}_1 + M_2\mathbf{v}_2)_{\text{initial}} = (M_1\mathbf{v}_1 + M_2\mathbf{v}_2)_{\text{final}}. \tag{25}$$

At any instant the sum of the momenta $M_1\mathbf{v}_1 + M_2\mathbf{v}_2$ of the two colliding particles is conserved, provided the mutual force is Newtonian.



If two moving point charges q_1 and q_2 pass close to one another, their orbits are deviated from initial straight-line paths. Momentum is conserved in this collision, but the forces are not Newtonian. The deviation from Newtonian behavior is important when the particles move very rapidly.

The simple Newtonian assumption about the forces is often not satisfied exactly, however. But the law of conservation of momentum is exact. In reality the interaction of 1 with 2 cannot be instantaneous, because forces do not propagate with infinite velocity, but at most with the speed of light (Chap. 10). Thus at any instant $\mathbf{F}_{12}$ is not exactly equal to $-\mathbf{F}_{21}$. It is only after the particles have achieved their final momenta that we expect to have the sum of the final momenta equal to the sum of the initial momenta. It is distinctly unsatisfying to base the conservation of linear momentum on the special assumption of Newtonian forces.

For Newtonian forces linear momentum is strictly conserved at every instant of time, even during the collision process itself. But for physical forces which are not instantaneous there will be stages during the intimate interaction of two particles when linear momentum is not conserved.

Second derivation. In this derivation we assume the validity of Galilean invariance and also the conservation of *energy and mass.* We consider two free particles 1 and 2 which initially have velocities $\mathbf{v}_1$, $\mathbf{v}_2$. The initial (and final) positions are assumed to be widely separated, so that at the initial and final epochs the particles do not interact. From earlier physics (or from Chap. 5) the initial kinetic energy of the particles is known to be

$$\tfrac{1}{2}M_1v_1^2 + \tfrac{1}{2}M_2v_2^2. \tag{26}$$

Now let the particles collide; it is *not* necessary that the collision be elastic. We shall find that momentum is conserved even if the collision is inelastic. The kinetic energy after the collision is

$$\tfrac{1}{2}M_1w_1^2 + \tfrac{1}{2}M_2w_2^2, \tag{27}$$

where $\mathbf{w}_1$ and $\mathbf{w}_2$ are the velocities after the collision, long enough after so that the particles no longer interact. The law of conservation of energy tells us that

$$\tfrac{1}{2}M_1v_1^2 + \tfrac{1}{2}M_2v_2^2 = \tfrac{1}{2}M_1w_1^2 + \tfrac{1}{2}M_2w_2^2 + \Delta\epsilon, \tag{28}$$

where $\Delta\epsilon$ (which can be either positive or negative) is the change in internal excitation energy of the particles, consequent to the collision. The internal excitation might be a rotation or an internal vibration; it might be the excitation of a bound electron from a low energy state to a high energy state. In an elastic collision $\Delta\epsilon = 0$, but we need not restrict the derivation to elastic collisions.† We have assumed here that the masses M_1, M_2 of the particles are unchanged in the collision.

Now view the same collision from the primed reference frame moving with the uniform velocity $\mathbf{V}$ with respect to the original unprimed frame. In the primed reference frame the initial velocities are $\mathbf{v}_1'$, $\mathbf{v}_2'$, and the final velocities are $\mathbf{w}_1'$, $\mathbf{w}_2'$. We have

$$\begin{aligned} \mathbf{v}_1' &= \mathbf{v}_1 - \mathbf{V}; \qquad & \mathbf{v}_2' &= \mathbf{v}_2 - \mathbf{V}; \\ \mathbf{w}_1' &= \mathbf{w}_1 - \mathbf{V}; \qquad & \mathbf{w}_2' &= \mathbf{w}_2 - \mathbf{V}. \end{aligned} \tag{29}$$

The statement of the law of conservation of energy in the primed frame is

$$\tfrac{1}{2}M_1(v_1')^2 + \tfrac{1}{2}M_2(v_2')^2 = \tfrac{1}{2}M_1(\mathbf{w}_1')^2 + \tfrac{1}{2}M_2(\mathbf{w}_2')^2 + \Delta\epsilon. \tag{30}$$

We have assumed that the excitation energy $\Delta\epsilon$ is unchanged on changing reference frames. This is an agreement with experiment.

If the law of conservation of energy is to be invariant under a Galilean transformation, then in both the primed and unprimed systems the initial kinetic energy must be equal to the final kinetic energy plus $\Delta\epsilon$, the internal excitation energy. That is, both (28) and (30) must hold. The law of conservation of energy in the primed system can also be expressed by substituting the transformation (29) in (30) and noting that $(v_1')^2 = v_1^2 - 2\mathbf{v}_1 \cdot \mathbf{V} + V^2$, etc., so that (30) becomes

$$\begin{aligned} \tfrac{1}{2}M_1(v_1^2 - 2\mathbf{v}_1 \cdot \mathbf{V} + V^2) + \tfrac{1}{2}M_2(v_2^2 - 2\mathbf{v}_2 \cdot \mathbf{V} + V^2) \\ = \tfrac{1}{2}M_1(w_1^2 - 2\mathbf{w}_1 \cdot \mathbf{V} + V^2) \\ + \tfrac{1}{2}M_2(w_2^2 - 2\mathbf{w}_2 \cdot \mathbf{V} + V^2) + \Delta\epsilon. \end{aligned} \tag{31}$$

Notice that the terms in V^2 cancel between the right and left sides. This expression is identical with the law of energy conservation (28) in the unprimed system, provided the scalar products cancel in (31):

$$(M_1\mathbf{v}_1 + M_2\mathbf{v}_2) \cdot \mathbf{V} = (M_1\mathbf{w}_1 + M_2\mathbf{w}_2) \cdot \mathbf{V}. \tag{32}$$

---

† In an inelastic collision there is no violation of the principle of conservation of energy. What happens is that kinetic energy lost or gained from the motion of the bodies appears as rotational, vibrational, or other excitational motion of the interior of the bodies. Such internal motion may often be called *heat* or thermal motion (Vol. V).

Equation (32) must hold for any value of $\mathbf{V}$. Hence the general solution of (32) is

$$\boxed{M_1\mathbf{v}_1 + M_2\mathbf{v}_2 = M_1\mathbf{w}_1 + M_2\mathbf{w}_2.} \tag{33}$$

This is precisely the *law of conservation of linear momentum.*

To review what we have done: We have assumed the conservation of energy and conservation of mass in a collision, and we have further assumed that these laws are valid in any inertial reference frame. That is, we assumed Galilean invariance. We found that the laws can be valid in different inertial frames *only* if linear momentum is conserved in the collision. We have not used the law of conservation of mass in its fullest generality. If the collision involved an exchange of some mass, so that after the collision M_1 became $\overline{M}_1$ and M_2 became $\overline{M}_2$, but with $M_1 + M_2 = \overline{M}_1 + \overline{M}_2$, the same steps used in the above derivation could be used to derive the conservation of momentum. This is demonstrated in the following example.

Chemical Reactions. We show that momentum is conserved in a chemical reaction in which the atoms of the reactants are rearranged or exchanged, while conserving the total mass. We assume that there are no external forces.

Let the reaction be represented by

$$\mathrm{A} + \mathrm{BC} \rightarrow \mathrm{B} + \mathrm{AC}, \tag{34}$$

where BC means a molecule consisting of atom B and atom C. In the reaction, atom C attaches itself to atom A to form AC.

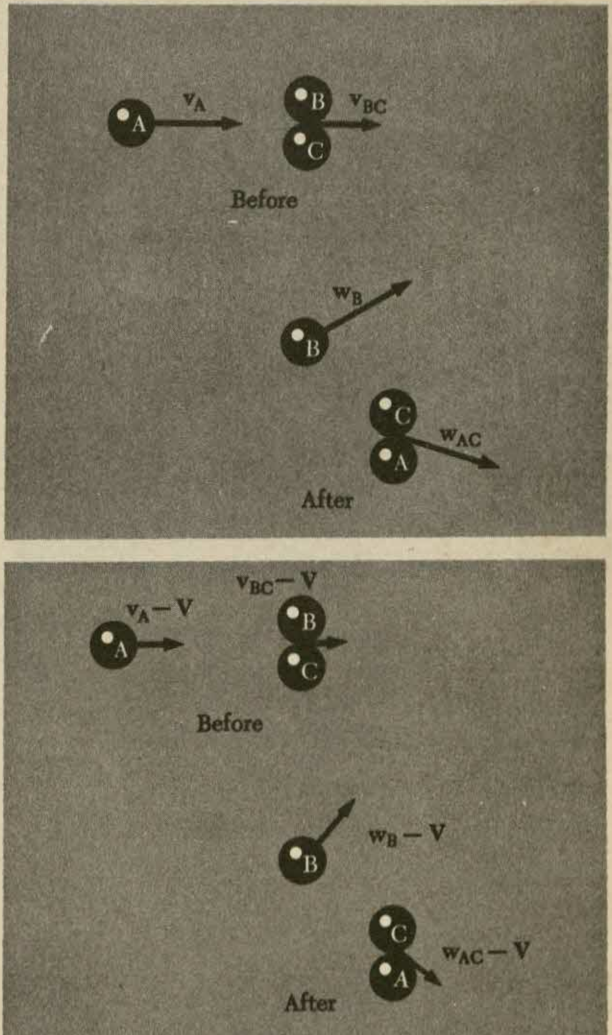
In one inertial frame the law of conservation of energy may be written as

$$\tfrac{1}{2}M_{\mathrm{A}}v_{\mathrm{A}}^2 + \tfrac{1}{2}(M_{\mathrm{B}} + M_{\mathrm{C}})v_{\mathrm{BC}}^2 = \tfrac{1}{2}M_{\mathrm{B}}w_{\mathrm{B}}^2 + \tfrac{1}{2}(M_{\mathrm{A}} + M_{\mathrm{C}})w_{\mathrm{AC}}^2 + \Delta\epsilon. \tag{35}$$

Here $\Delta\epsilon$ represents changes in the binding energy of the molecules taking part in the reaction. In a second inertial frame moving with velocity $\mathbf{V}$ with respect to the first, the law of conservation of energy may be written as, with $\mathbf{v}_{\mathrm{A}}$ replaced by $\mathbf{v}_{\mathrm{A}} - \mathbf{V}$, etc.,

$$\tfrac{1}{2}M_{\mathrm{A}}(\mathbf{v}_{\mathrm{A}} - \mathbf{V})^2 + \tfrac{1}{2}(M_{\mathrm{B}} + M_{\mathrm{C}})(\mathbf{v}_{\mathrm{BC}} - \mathbf{V})^2 = \tfrac{1}{2}M_{\mathrm{B}}(\mathbf{w}_{\mathrm{B}} - \mathbf{V})^2 + \tfrac{1}{2}(M_{\mathrm{A}} + M_{\mathrm{C}})(\mathbf{w}_{\mathrm{AC}} - \mathbf{V})^2 + \Delta\epsilon. \tag{36}$$

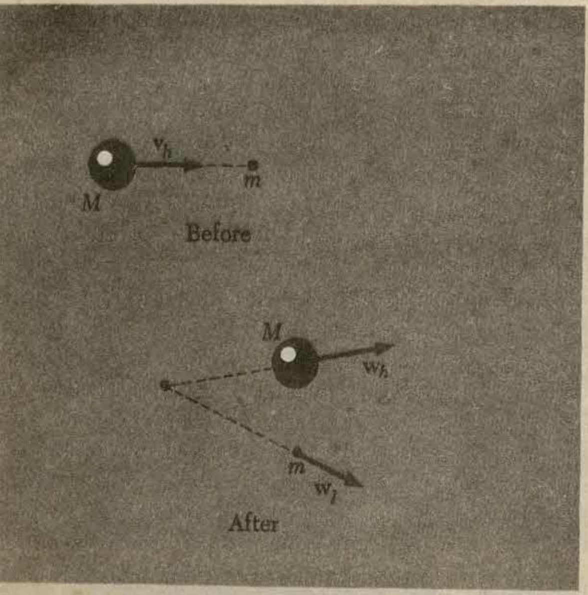
On writing out the squares of the quantities in parentheses we see that Eqs. (35) and (36) are consistent if



A collision between atom A and molecule BC resulting in atom B and molecule AC. The collision is viewed from two different reference frames.

$$M_A\mathbf{v}_A + (M_B + M_C)\mathbf{v}_{BC} = M_B\mathbf{w}_B + (M_A + M_C)\mathbf{w}_{AC}, \quad (37)$$

which is exactly a statement of the law of conservation of linear momentum.



Example. Collision of heavy particle with light particle. A heavy particle of mass M collides elastically with a light particle of mass m. The light particle is initially at rest. The initial velocity of the heavy particle is $\mathbf{v}_h = v_h\hat{\mathbf{x}}$; the final velocity is $\mathbf{w}_h$. If the particular collision is such that the light particle goes off in the forward $(+\hat{\mathbf{x}})$ direction, what is its velocity $\mathbf{w}_l$? What fraction of the energy of the heavy particle is lost in this collision?

From momentum conservation there can be no $\hat{\mathbf{y}}$ component to the final velocity of the heavy particle in this particular collision, so that

$$Mv_h\hat{\mathbf{x}} = Mw_h\hat{\mathbf{x}} + mw_l\hat{\mathbf{x}}, \quad (38)$$

or

$$Mv_h = Mw_h + mw_l. \quad (39)$$

From energy conservation we have (with $\Delta\epsilon = 0$ for an elastic collision)

$$\tfrac{1}{2}Mv_h^2 = \tfrac{1}{2}Mw_h^2 + \tfrac{1}{2}mw_l^2. \quad (40)$$

This may be written with the help of (39) as

$$\frac{1}{2}M\left(w_h^2 + \frac{2m}{M}w_hw_l + \frac{m^2}{M^2}w_l^2\right) = \frac{1}{2}Mw_h^2 + \frac{1}{2}mw_l^2. \quad (41)$$

If $m \ll M$, it is convenient to neglect the term of order m^2/M^2, so that (41) reduces to

$$mw_hw_l \cong \tfrac{1}{2}mw_l^2; \quad (42)$$

or

$$w_l \cong 2w_h. \quad (43)$$

Thus the light particle goes off at approximately twice the velocity of the heavy particle. It follows further, from (43) substituted in (39), that

$$Mv_h \cong Mw_h + 2mw_h; \quad (44)$$

or

$$\frac{\Delta v_h}{v_h} \cong \frac{v_h - w_h}{w_h} \cong \frac{2m}{M}. \quad (45)$$

The fractional energy loss of the heavy particle is

$$\frac{\Delta(\frac{1}{2}Mv_h^2)}{\frac{1}{2}Mv_h^2} = \frac{Mv_h\Delta v_h}{\frac{1}{2}Mv_h^2} = \frac{2\Delta v_h}{v_h} \cong \frac{4m}{M}, \tag{46}$$

using (45).

Other examples of the application of the conservation of linear momentum are treated in Chap. 6.

Fictitious Forces. We know from Newton's second law that in an inertial frame of reference

$$\mathbf{F} = M\mathbf{a}_I, \tag{47}$$

in which the left-hand side is the applied force, and $\mathbf{a}_I$ is the acceleration as observed in the inertial frame. The mass M is assumed to be constant. The subscript I has been added to $\mathbf{a}$ to emphasize the word *inertial.* In a noninertial frame, such as on the rotating earth, we know that (47) is not valid as it stands. The reason is that an acceleration $\mathbf{a}_0$ has been left out which should have been included, namely the acceleration of the noninertial frame relative to the inertial frame. If $\mathbf{a}$ is the acceleration of a body as measured in the noninertial frame, we have $\mathbf{a} + \mathbf{a}_0 = \mathbf{a}_I$, or

$$\boxed{\mathbf{F} = M(\mathbf{a} + \mathbf{a}_0).} \tag{48}$$

If we do experiments in a noninertial frame, we must always be sure to include $\mathbf{a}_0$ *in the force equation.* In working in a noninertial frame it is often convenient to think in terms of a quantity $\mathbf{F}_0$ such that (48) appears as

$$\boxed{\mathbf{F} + \mathbf{F}_0 = M\mathbf{a},} \tag{49}$$

where

$$\boxed{\mathbf{F}_0 \equiv -M\mathbf{a}_0} \tag{50}$$

is called the *fictitious force* or *pseudoforce.* The fictitious force as defined is equal but opposite to the mass times the acceleration of the noninerital frame; it simply represents the effect of the acceleration of the noninertial frame itself; it is the quantity we have to add to the true force $\mathbf{F}$ to make the sum equal to $M\mathbf{a}$ *as observed in the noninertial frame.* Anything fictitious in physics tends to seem confusing; you can always resolve any problem by going back to Eq. (48).

Example. Accelerometer. Suppose the force applied to a mass M by a stretched spring in the x direction is $F_x = -Cx$, where C is a constant. Consider a noninertial frame with the acceleration $\mathbf{a}_0 = a_0\hat{\mathbf{x}}$ in the x direction. If the spring is at rest in this noninertial frame, then its acceleration $\mathbf{a}$ in this frame is zero, and $\mathbf{F} = M(\mathbf{a} + \mathbf{a}_0)$ reduces to $F_x = Ma_0$, and $\mathbf{F} + \mathbf{F}_0 = M\mathbf{a}$ reduces to

$$F_x + F_{0x} = 0. \tag{51}$$

Thus

$$-Cx + F_{0x} = 0; \qquad -Cx - Ma_0 = 0, \tag{52}$$

or

$$x = -\frac{Ma_0}{C}. \tag{53}$$

The displacement x is proportional to the acceleration a_0 of the noninertial frame. The noninertial frame might be an aircraft or an automobile. We see that (53) describes the operation of an *accelerometer* in which a mass M is attached to a spring and constrained to move in one dimension. The displacement x of the mass measures the acceleration a_0 of the noninertial reference frame.

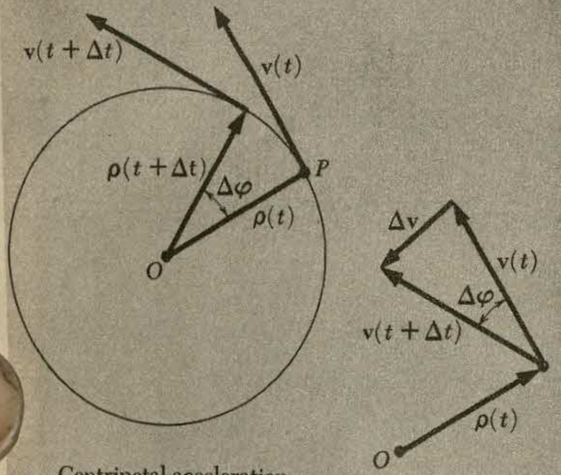
Example. Centrifugal force and centripetal acceleration in a uniformly rotating frame. Although we will almost immediately discuss rotating frames in detail, it is worthwhile to discuss a simple and common example straightaway. Consider a point mass M at rest in a noninertial frame, so that in this frame $\mathbf{a} = 0$. The noninertial frame rotates uniformly about an axis fixed with respect to an inertial frame. The acceleration of the point in question was seen in Chap. 2 to be

$$\mathbf{a}_0 = -\omega^2\boldsymbol{\rho} \tag{54}$$

with respect to the inertial frame, where $\boldsymbol{\rho}$ is directed outward to the particle from the axis. The vector $\boldsymbol{\rho}$ is taken perpendicular to the axis. Here we use $\boldsymbol{\rho}$, the Greek letter rho, to denote the two-dimensional analog of the three-dimensional position vector $\mathbf{r}$.

Equation (54) is the famous *centripetal acceleration.* The mass might be constrained to be at rest by a stretched spring. The specification that in the noninertial frame $\mathbf{a} = 0$ leads to, by Eq. (49),

$$\mathbf{F} = -\mathbf{F}_0 = M\mathbf{a}_0 = -M\omega^2\boldsymbol{\rho}. \tag{55}$$



Centripetal acceleration

If point P describes circular motion with radius ρ and with constant speed v, the acceleration $\mathbf{a}$ of P is always directed toward the center O. The magnitude of $\mathbf{a}$ is $a = v^2/\rho = \omega^2\rho$, where $\omega = d\varphi/dt = v/\rho$. To see this, note that $\mathbf{a} = \lim_{\Delta t \to 0} \frac{\Delta \mathbf{v}}{\Delta t}$ is directed toward O always. Also

$$a = \lim_{\Delta t \to 0} \left|\frac{\Delta \mathbf{v}}{\Delta t}\right| = \lim_{\Delta t \to 0} \frac{v \sin \Delta\varphi}{\Delta t} = \lim_{\Delta t \to 0} v\frac{\Delta\varphi}{\Delta t}.$$

Thus $a = v\omega = v^2/\rho = \omega^2\rho$. This result was also derived analytically in Chap. 2.

The fictitious force $\mathbf{F}_0$ in this example is called the *centrifugal force;* it is $\mathbf{F}_0 = M\omega^2\boldsymbol{\rho}$; and it is directed away from the axis. The centrifugal force is balanced in this example by the elastic force $\mathbf{F}$ of the spring, in order to produce zero acceleration (mass at rest) in the rotating noninertial frame.

If $M = 100$ gm, $\rho = 10$ cm, and the frame rotates at 100 revolutions per second, what is the value of the centrifugal force? We have $F_0 = M\omega^2\rho = (10^2)(2\pi \cdot 100)^2(10) \approx 4 \times 10^8$ dynes.

Example. Experiments in a freely falling elevator. Let the acceleration of a noninertial frame (a freely falling elevator) be

$$\mathbf{a}_0 = -g\hat{\mathbf{z}}, \tag{56}$$

where $\hat{\mathbf{z}}$ is measured upward from the surface of the earth and g is the acceleration of gravity. This acceleration corresponds to free fall under gravity. From (50) the fictitious force on a mass M in the noninertial frame is

$$\mathbf{F}_0 = -M\mathbf{a}_0 = Mg\hat{\mathbf{z}}. \tag{57}$$

An unattached body in the elevator is acted on by the sum of the gravitational force $\mathbf{F} = -Mg\hat{\mathbf{z}}$ and the fictitious force $\mathbf{F}_0 = Mg\hat{\mathbf{z}}$, so that the total apparent force in the noninertial frame of the freely falling elevator is zero:

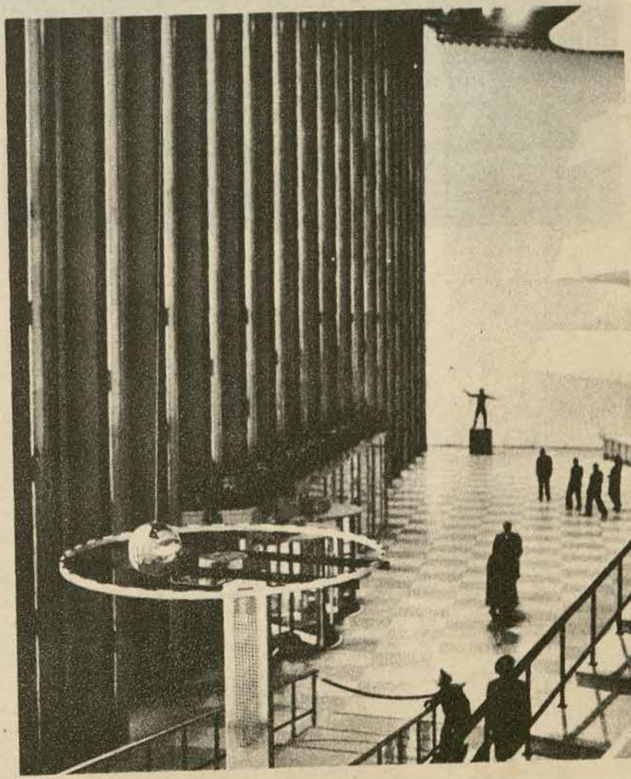
$$\mathbf{F} + \mathbf{F}_0 = 0. \tag{58}$$

Thus the body is unaccelerated in the noninertial frame. This is a form of "weightlessness." The body appears to remain suspended in space, if it has no initial velocity relative to the elevator.

Example. Foucault pendulum. The Foucault pendulum is a device which illustrates the rotation of the earth; it demonstrates the fact that the earth is not an inertial reference frame. The experiment was first performed publicly by Foucault in 1851, under the great dome of the Pantheon in Paris, using a 28-kg mass on a wire suspension nearly 70 m long. The attachment of the upper end of the wire allows the pendulum to swing with equal freedom in any direction. The period (see Chap. 7) of a pendulum of this length is about 17 sec.

Around the point on the floor directly under the point of suspension there was constructed a circular railing about 3 m

Foucault pendulum as installed in the United Nations Headquarters in New York. The sphere, seen at left, is gold-plated and weighs 200 pounds. It is suspended from the ceiling 75 feet above the floor of the lobby. A stainless steel wire holds it in such a manner as to allow it to swing freely in any plane. The sphere swings directly over the raised metal ring, which is about six feet in diameter. The sphere swings continuously as a pendulum, its plane shifting slowly in a clockwise direction, thus offering visual proof of the rotation of the earth. A complete cycle takes approximately 36 hours and 45 minutes. Inscribed on it is a message by Queen Juliana of The Netherlands: "It is a privilege to live today and tomorrow." (*United Nations photograph*)



in radius. On this railing was piled a ridge of sand so that a metal point extending downward from the pendulum brushed aside the sand at each swing.

With successive swings it appeared that the plane of the motion of the pendulum moved in a clockwise direction as viewed from above. In an hour the pendulum changed the plane of its swing by over 11 degrees. A full circuit was completed in about 32 hours. In one swing the plane moved by 3 mm, as measured at the circle of sand.

Why does the plane of the pendulum rotate? If the Foucault experiment were carried out at the North Pole of the earth, we could see immediately that the plane of motion of the pendulum would remain fixed in an inertial frame while the earth rotates under the pendulum once every 24 hours. The rotation of the earth is counterclockwise as viewed from above the North Pole (say from Polaris), so to an observer on a ladder on the earth at the North Pole the plane of the pendulum appears to rotate clockwise relative to him.

The situation appears different (and more difficult to analyze) when we leave the North Pole, and the time for a full circuit is longer. Consider the relative velocities of the extreme north and south points of the circle of sand of radius r. The south point is farther from the axis of rotation of the earth and will therefore move in space faster than the north point. If ω denotes the angular velocity of the earth and R denotes the radius of the earth, then the center of the circle of sand moves at velocity $\omega R \cos \varphi$, where φ is the latitude of Paris (48°51′N) as measured from the equator of the earth. The northernmost point on the ring moves at velocity

$$v_N = \omega R \cos \varphi - \omega r \sin \varphi, \tag{58a}$$

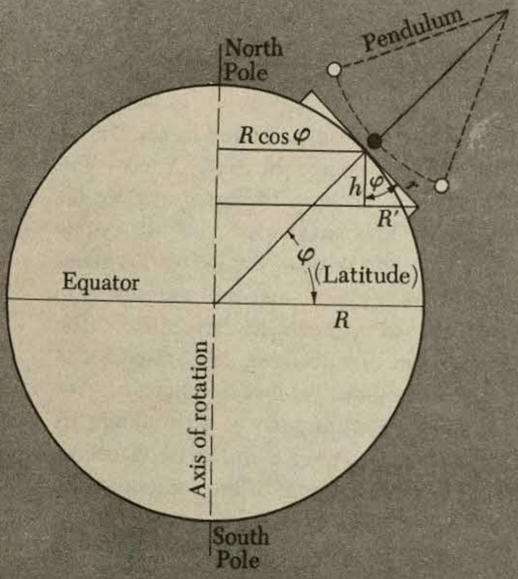
as we see from the figure, and the southernmost point moves at velocity

$$v_S = \omega R \cos \varphi + \omega r \sin \varphi. \tag{58b}$$

The difference between either velocity and that of the center of the ring is

$$\Delta v = \omega r \sin \varphi. \tag{58c}$$

If the pendulum is started in the north-south plane by a push from rest at the center of the ring, the east-west component of velocity in space will be the same as that of the center of the ring.



The Foucault pendulum, greatly exaggerated in size in relation to the earth, is shown at approximately the angle of latitude φ of Paris. The circle of sand beneath the pendulum has a radius r. The distance from the earth's axis to the center of the pendulum's swing is $R \cos \varphi$. Because of the earth's rotation the south side of the sand moves faster than the north side (relative to an inertial frame).

The circumference of the ring is $2\pi r$, so that the time T_0 for a full circuit is, if Δv is constant around the ring,

$$T_0 = \frac{2\pi r}{\omega r \sin\varphi} = \frac{24 \text{ hours}}{\sin\varphi}. \tag{58d}$$

At the equator $\sin\varphi = 0$ and the time becomes infinite.

What happens when the plane of the pendulum reaches the east-west plane through the center of the ring? Why should Δv remain the same here as in the north-south plane? This is hard to see without reference to a globe. Take a piece of cardboard or stiff paper and hold it out from a globe. Let it nearly touch the globe at Paris; let it be normal to the globe at this point and lie in an east-west plane. The normal direction to the surface of the globe is the line of the pendulum wire. With one hand hold the plane of the cardboard fixed while rotating the globe slowly with the other hand. Notice that one side of the line of near contact of the cardboard and the globe appears to move southward and the other side appears to move northward. Contemplation or detailed analysis gives the same value of Δv as found above: the plane of the pendulum actually turns relative to the ring on the Pantheon floor with the constant angular velocity $\omega \sin\varphi$, where ω is the angular velocity of the earth and φ is the latitude. Many mechanics texts at the junior level have a mathematical treatment of the equation of motion of a Foucault pendulum.

Newtonian Law of Universal Gravitation. It is convenient to prepare for the following chapters with a review of the law of gravitation. This law states that every mass M_1 attracts every other mass M_2 in the universe with a force

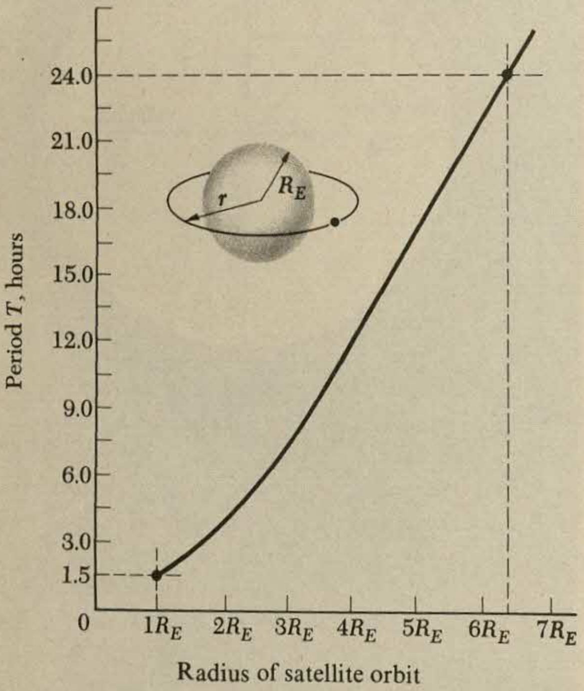
$$\mathbf{F} = -\frac{GM_1M_2}{r^2}\hat{\mathbf{r}}, \tag{59}$$

where $\mathbf{r}$ is the vector from M_1 to M_2 and G is a constant having the value

$$6.67 \times 10^{-8} \text{ dyne-cm}^2/\text{gm}^2 \quad \text{or} \quad 6.67 \times 10^{-11} \text{ newtons-m}^2/\text{kg}^2.$$

The gravitational force is a central force: the force is directed along the line connecting the two point masses. The determination of the value of G is usually treated in high school texts. The classical experiment is that of Cavendish.

It is also known experimentally to high accuracy that the gravitational and inertial masses of a body are equal. This will be discussed in Chap. 14. What is meant is that the value of the mass M to be used in the gravitational force equation above is equal to the value of the mass of the same body used in Newton's second law $\mathbf{F} = M\mathbf{a}$. The mass in the gravitational equation is called the *gravitational mass,* and the mass in Newton's second law is called the *inertial mass.* The classical experiments on the equality of the two masses were carried out by Eötvös; a recent experiment is described by R. H. Dicke [*Sci. American* **205,** 84 (December, 1961)]. The Eötvös experiment is described in Chap. 14.



The period of an artificial satellite describing a circular orbit about the earth. This graph was obtained by plotting (61) in terms of the period $T = 2\pi/\omega$.

Example. Satellite in circular orbit. Consider a satellite in a circular orbit concentric and coplanar with the equator of the earth. At what radius r of the orbit will the satellite appear to remain stationary when viewed by observers fixed on the earth? We suppose the sense of rotation of the orbit is the same as that of the earth.

In a circular orbit the gravitational attraction is equal and opposite to the centrifugal force:

$$\frac{GM_eM_s}{r^2} = M_s\omega^2 r, \qquad (60)$$

where M_e is the mass of the earth and M_s the mass of the satellite. We rearrange (60) as

$$r^3 = \frac{GM_e}{\omega^2} = \frac{GM_eT^2}{(2\pi)^2}, \qquad (61)$$

where T is the period. In our special problem we want ω for the satellite orbit to equal the angular frequency ω_e of the earth about its axis. The angular frequency of the earth is

$$\omega_e = \frac{2\pi}{8.64 \times 10^4} = 7.3 \times 10^{-5}\ \text{sec}^{-1}, \qquad (62)$$

whence (61) becomes, with $\omega = \omega_e$,

$$r^3 \approx \frac{(6.67 \times 10^{-8})(5.98 \times 10^{27})}{(7.3 \times 10^{-5})^2} \approx 75 \times 10^{27}, \qquad (63)$$

or

$$r \approx 4.2 \times 10^9\ \text{cm}. \qquad (64)$$

The radius of the earth is 6.38×10^8 cm. The distance (64) is roughly one-tenth of the distance to the moon.

Further reading

PSSC, chaps. 20, 21, 23, and 24 (paragraph 7).

Collier's Encyclopedia (1964); an excellent elementary discussion of the Foucault pendulum is given under Foucault.

Film list

"Frames of Reterence" (28 min) Hume and Ivey (PSSC-MLA 0307). An excellent film whose treatment of frames of reference parallels closely the material in the first part of this chapter and gives many graphic examples of the ideas involved.

"Universal Gravitation" (31 min) Hume and Ivey (PSSC-MLA 0309). A historical review of the developments leading up to and including Newton's discovery of the law of universal gravitation.

"Forces" (23 min) J. Zacharias (PSSC-MLA). Gives a clear demonstration of a large-scale torsion balance used to measure the gravitational constant in a paraphrase of Cavendish's experiment.

Problems

1. *Velocity of earth about sun.* (*a*) What is the velocity of the center of the earth in its orbit about the sun? (Treat the orbit as circular. The radius of the orbit is listed in the table of constants inside the cover of this volume.) *Ans.* 3.0×10^6 cm/sec.

(*b*) What is the ratio of this velocity to the velocity of light? *Ans.* 10^{-4}.

2. *Velocity due to rotation.* What is the velocity of a point on the surface of the earth at the equator relative to the center of the earth? *Ans.* 4.7×10^4 cm/sec.

3. *Falling body.* A massive body is dropped from a height of 100 m. How long does it take to reach the ground?

4. *Acceleration in circular motion.* (*a*) What is the centripetal acceleration in cm/sec^2 of a body of mass $M = 1$ kg ($= 1000$ gm) moving in a circle of radius $R = 100$ cm at an angular velocity $\omega = 10$ sec^{-1}? *Ans.* 1×10^4 cm/sec^2.

(*b*) What is the centrifugal force in dynes? *Ans.* 1×10^7 dynes.

5. *Angular frequency and period.* A tuning fork vibrates at a frequency f of 60 cycles per second.

(*a*) What is the angular frequency ω, in radians per second? Note that we can speak of an angular frequency in connection with any periodic oscillation and not only in connection with motion in a circle.

(*b*) What is the period T of the motion, in seconds? (The period is the time of a complete cycle of the motion.)

6. *Acceleration of the earth.* Suppose that an opaque cloud layer covered the earth. Describe an experiment which would determine unambiguously (with respect to both rotational and translational accelerations) whether the earth is an inertial frame.

7. *Trajectory, free motion.* An object is projected horizontally at $t = 0$ with a speed of 1000 cm/sec, from a frame which undergoes an upward acceleration of 300 cm/sec^2, starting from rest at $t = 0$ on the ground.

(*a*) What is the trajectory, expressed as $x = f(t)$; $y = F(t)$, of the object relative to an inertial frame with origin at rest at the starting point on the ground? (Neglect the rotation of the earth.)

(*b*) Sketch the path of the object as seen from a frame of reference which undergoes an upward acceleration of 300 cm/sec^2, starting from rest at $t = 0$ on the ground, and also from the inertial frame referred to in (*a*).

8. *Acceleration in circular motion.* An object moves in a circular path with a constant speed v of 50 cm/sec. The velocity vector $\mathbf{v}$ changes direction by 30° in 2 sec.

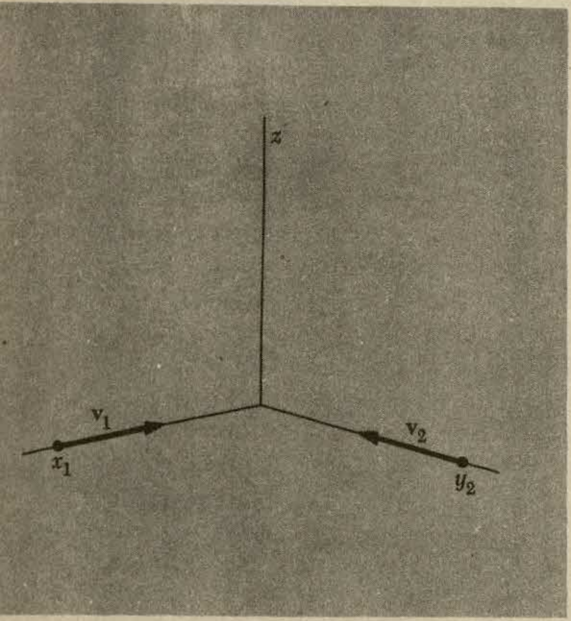
(*a*) Find the magnitude of the change in velocity, $\Delta\mathbf{v}$.

(*b*) Find the magnitude of the average acceleration during the interval.

(*c*) What is the centripetal acceleration of the uniform circular motion?

Ans. 13.1 cm/sec^2.

9. *Effective force due to rotation.* An object fixed with respect to the surface of a planet identical in mass and radius to the earth experiences zero gravitational acceleration at the equator. What is the length of a day on that planet? *Ans.* 1.3 hr.



10. *Collision courses.* Initially two particles are at positions $x_1 = 5$ cm, $y_1 = 0$; and $x_2 = 0$, $y_2 = 10$ cm, with $\mathbf{v}_1 = -4 \times 10^4\ \hat{\mathbf{x}}$ cm/sec and $\mathbf{v}_2$ is along $-\hat{\mathbf{y}}$ as in the figure.

(*a*) What must be the value of $\mathbf{v}_2$ if they are to collide?

Ans. $-8 \times 10^4\ \hat{\mathbf{y}}$ cm/sec.

(*b*) What is the value of $\mathbf{v}_r$, the relative velocity?

Ans. $4 \times 10^4\ (2\hat{\mathbf{y}} - \hat{\mathbf{x}})$ cm/sec.

(*c*) Establish a general criterion for recognizing a collision course for two objects, in terms of their positions $\mathbf{r}_1$, $\mathbf{r}_2$ and velocities $\mathbf{v}_1$, $\mathbf{v}_2$.

11. *Collision kinematics.* Two masses constrained to move in a horizontal plane collide. Given initially that $M_1 = 85$ gm, $M_2 = 200$ gm, $\mathbf{v}_1 = 6.4\hat{\mathbf{x}}$ cm/sec, $\mathbf{v}_2 = -6.7\hat{\mathbf{x}} - 2.0\hat{\mathbf{y}}$ cm/sec.

(*a*) Find the velocity of the center of mass. The position of the center of mass is defined as $\mathbf{R}_{\text{c.m.}} = (M_1\mathbf{r}_1 + M_2\mathbf{r}_2)/(M_1 + M_2)$, so that the velocity of the center of mass is given by $\dot{\mathbf{R}}_{\text{c.m.}} = \dfrac{M_1\mathbf{v}_1 + M_2\mathbf{v}_2}{M_1 + M_2}$.

Ans. $-2.8\hat{\mathbf{x}} - 1.4\hat{\mathbf{y}}$ cm/sec.

(*b*) Find the total linear momentum. *Ans.* $-796\hat{\mathbf{x}} - 400\hat{\mathbf{y}}$ gm-cm/sec.

(*c*) Find the velocities in the reference frame in which the center of mass is at rest.

Ans. $\mathbf{v}_1' = 9.2\hat{\mathbf{x}} + 1.4\hat{\mathbf{y}}$ cm/sec; $\mathbf{v}_2' = -3.9\hat{\mathbf{x}} - 0.6\hat{\mathbf{y}}$ cm/sec.

After collision $|\mathbf{w}_1| = 9.2$ cm/sec, $\mathbf{w}_2 = -4.4\hat{\mathbf{x}} + 1.9\hat{\mathbf{y}}$ cm/sec.

(*d*) What is the direction of $\mathbf{w}_1$? *Ans.* −84° with respect to x axis.

(*e*) What is the value of the relative velocity $\mathbf{w}_r = \mathbf{w}_1 - \mathbf{w}_2$?

Ans. $5.4\hat{\mathbf{x}} - 11\hat{\mathbf{y}}$ cm/sec.

(*f*) What are the initial and final total kinetic energies in the laboratory and center-of-mass frame? Is the collision elastic or inelastic?

12. *Inelastic collision.* Two objects ($M_1 = 2$ gm; $M_2 = 5$ gm) possess velocities $\mathbf{v}_1 = 10\hat{\mathbf{x}}$ cm/sec; $\mathbf{v}_2 = 3\hat{\mathbf{x}} + 5\hat{\mathbf{y}}$ cm/sec just prior to a collision during which they become permanently attached to each other.

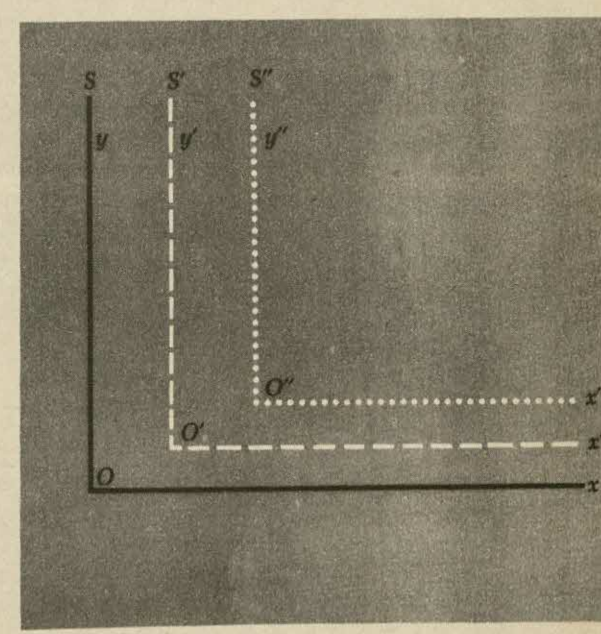
(*a*) What is the velocity of the center of mass, as defined in Prob. 11?

(*b*) What is the final momentum of the combination in the laboratory system?

(*c*) What is the final momentum in the center-of-mass system?

(*d*) What fraction of the initial total kinetic energy is associated with the motion after collision? *Ans.* 0.72.

13. *Motion in noninertial reference frames.* Two noninertial reference frames, S′ and S″, coincide at $t = 0$ with an inertial frame S. S″ has an initial velocity v_0 along the x axis while S′ is at rest. At $t = 0$ both frames S′ and S″ experience the same acceleration **a** along the x axis.



(*a*) What are the positions of O' and O'' with respect to O as functions of time?

(*b*) Relate the positions x',x'' of a particle in S′ and S″ to the position x in S.

(*c*) Write down the equation of motion of a particle experiencing a constant force **F** in S. Transform this equation to S′ and S″. Do any fictitious forces appear in the transformed equations?

(*d*) Are the applied and fictitious forces respectively the same in the two moving frames? What is the velocity of S″ with respect to S′?

(*e*) In general when two noninertial frames move with a constant relative velocity, what can be said about the real and fictitious forces in the two frames?

14. *Gravitational attraction, earth–moon.* Evaluate the gravitational force in dynes between the earth and the moon. *Ans.* 2×10^{25} dynes.

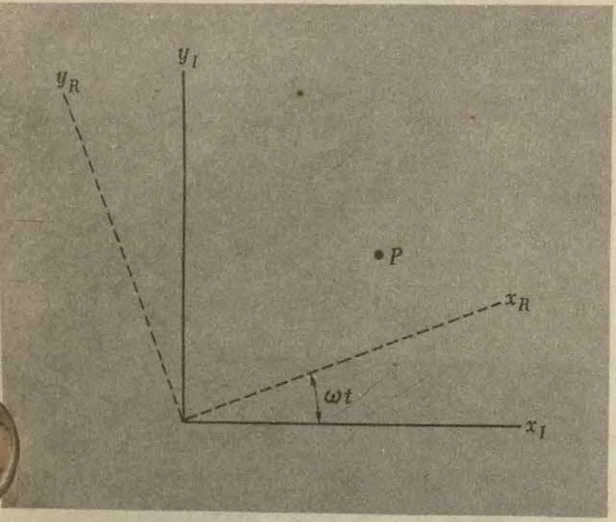
15. *Satellite orbit.* Consider a circular satellite orbit which lies just outside the equator of a homogeneous spherical planet of mass density ρ. Show that the period T of such an orbit depends only on the density of the planet. (Give the equation.)

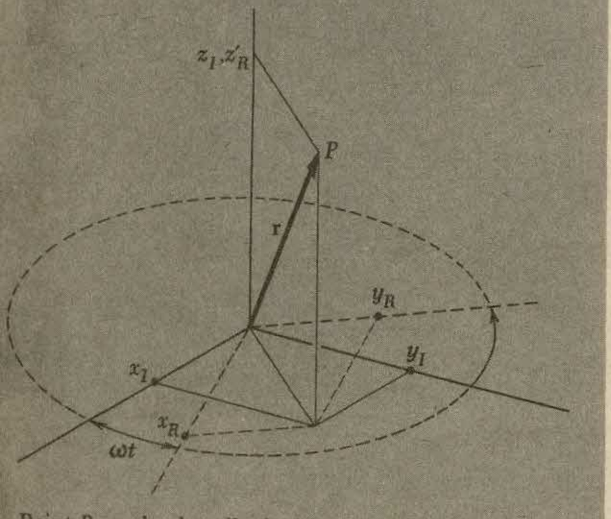
Advanced Topic. Velocity and acceleration in rotating coordinate systems†

We now consider a noninertial reference frame which is rotating with constant angular velocity ω about the z axis of an inertial frame. The importance of this problem lies in the fact that the earth is rotating; a reference frame fixed on the surface of the earth is not an inertial frame. We have to add terms to $\mathbf{F} = M\mathbf{a}$ to represent the acceleration of a reference frame fixed on the surface of the earth. Besides the centripetal acceleration, we

---

† This section and the five examples which immediately follow may be omitted from a minimum program.





Point P can be described with coordinates $x_I y_I z_I$ of the inertial frame or with coordinates $x_R y_R z_R$ of the rotating frame. The rotation is about the z axis.

$$x_I = x_R \cos \omega t - y_R \sin \omega t$$
$$y_I = x_R \sin \omega t + y_R \cos \omega t$$
$$z_I = z_R$$

pick up in the analysis the Coriolis acceleration, which is important in the large-scale motion of sea and air currents.

The coordinates (x_R, y_R, z_R) of a point P viewed from the rotating frame may be related simply to the coordinates (x_I, y_I, z_I) of the same point as viewed from the inertial frame. We see on studying the geometry in the accompanying figure that

$$x_I = x_R \cos \omega t - y_R \sin \omega t; \tag{65a}$$
$$y_I = x_R \sin \omega t + y_R \cos \omega t; \tag{65b}$$
$$z_I = z_R. \tag{65c}$$

The relations between the velocity components in the two frames are found on differentiating (65) with respect to time. (For compactness we sometimes use a dot over a quantity to denote differentiation with respect to the time. Thus $\dot{x} \equiv dx/dt \equiv v_x$, and $\ddot{x} \equiv d^2x/dt^2 \equiv \dot{v}_x \equiv dv_x/dt$.) We have

$$\dot{x}_I = \dot{x}_R \cos \omega t - \omega x_R \sin \omega t - \dot{y}_R \sin \omega t - \omega y_R \cos \omega t; \tag{66a}$$
$$\dot{y}_I = \dot{x}_R \sin \omega t + \omega x_R \cos \omega t + \dot{y}_R \cos \omega t - \omega y_R \sin \omega t; \tag{66b}$$
$$\dot{z}_I = \dot{z}_R. \tag{66c}$$

We have taken ω to be constant, in the interest of simplicity. Notice for a particle at rest $(\dot{x}_R = \dot{y}_R = \dot{z}_R = 0)$ in the rotating frame that (66) reduces to

$$\dot{x}_I = -\omega x_R \sin \omega t - \omega y_R \cos \omega t; \qquad \dot{y}_I = \omega x_R \cos \omega t - \omega y_R \sin \omega t. \tag{67}$$

Similarly, for a particle at rest in the inertial frame $(\dot{x}_I = \dot{y}_I = \dot{z}_I = 0)$, we have

$$\dot{x}_R - \omega y_R = 0; \qquad \dot{y}_R + \omega x_R = 0; \qquad \dot{z}_R = 0, \tag{68}$$

from (66).

The acceleration components are found on differentiating (66) with respect to the time:

$$\begin{aligned} \ddot{x}_I = \ddot{x}_R \cos \omega t &- 2\omega \dot{x}_R \sin \omega t - \omega^2 x_R \cos \omega t \\ &- \ddot{y}_R \sin \omega t - 2\omega \dot{y}_R \cos \omega t + \omega^2 y_R \sin \omega t; \end{aligned} \tag{69a}$$

$$\begin{aligned} \ddot{y}_I = \ddot{x}_R \sin \omega t &+ 2\omega \dot{x}_R \cos \omega t - \omega^2 x_R \sin \omega t \\ &+ \ddot{y}_R \cos \omega t - 2\omega \dot{y}_R \sin \omega t - \omega^2 y_R \cos \omega t; \end{aligned} \tag{69b}$$

$$\ddot{z}_I = \ddot{z}_R. \tag{69c}$$

We notice for a particle at rest in the rotating frame that, with the help of (65), (69) reduces to

$$\ddot{x}_I = -\omega^2(x_R \cos \omega t - y_R \sin \omega t) = -\omega^2 x_I; \tag{70a}$$
$$\ddot{y}_I = -\omega^2(x_R \sin \omega t + y_R \cos \omega t) = -\omega^2 y_I. \tag{70b}$$

Equation (70) may be written in vector form as

$$\mathbf{a}_I = -\omega^2 \boldsymbol{\rho}_I, \tag{71}$$

where $\mathbf{a}_I \equiv \ddot{\mathbf{r}}_I$ is the acceleration of the particle relative to the inertial frame, and $\boldsymbol{\rho}_I = x_I\hat{\mathbf{x}}_I + y_I\hat{\mathbf{y}}_I$, as in (54) earlier. Equation (71) is the expression of the usual centripetal acceleration.

The essential physical results are contained in (69). It is useful to rewrite (69) in vector form. After some tedious algebra (which we give below), we obtain the important result

$$\boxed{\underset{\substack{\text{Acceleration}\\ \text{in inertial}\\ \text{frame}}}{\mathbf{a}_I} = \underset{\substack{\text{Acceleration}\\ \text{in rotating}\\ \text{frame}}}{\mathbf{a}_R} + \underset{\substack{\text{Coriolis}\\ \text{acceleration}}}{2\boldsymbol{\omega} \times \mathbf{v}_R} + \underset{\substack{\text{Centripetal}\\ \text{acceleration}}}{\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}_R)},} \tag{72}$$

where quantities with a subscript R are as observed in the rotating frame. The angular frequency ω has been taken to be constant. We have labeled the two terms on the right with their conventional names.

We can confirm that (72) reduces to (69) if $\boldsymbol{\omega} = \omega\hat{\mathbf{z}}$; for the projection of *both sides* of (72) on the x_I axis we have

$$(\mathbf{a}_I)_{x_I} = (\mathbf{a}_R)_{x_I} + 2(\boldsymbol{\omega} \times \mathbf{v}_R)_{x_I} + [\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}_R)]_{x_I}. \tag{72a}$$

We have written this out explicitly to emphasize the projection on the x_I axis. For example, the projection of the vector $\mathbf{a}_R$ on the x_I axis is

$$(\mathbf{a}_R)_{x_I} = \ddot{x}_R \cos \omega t - \ddot{y}_R \sin \omega t, \tag{72b}$$

because any vector in the rotating frame (such as $\mathbf{a}_R$) must project onto x_I in the same way as any other vector. From (65*a*) we see how to project the vector $\mathbf{r}_R$ onto x_I, and we follow the same track when projecting $\mathbf{a}_R$ onto x_I. This leads to (72*b*). The next term in (72*a*) is

$$\begin{aligned} 2(\boldsymbol{\omega} \times \mathbf{v}_R)_{x_I} &= 2\omega_{y_I}(\mathbf{v}_R)_{z_I} - 2\omega_{z_I}(\mathbf{v}_R)_{y_I} \\ &= -2\omega(\dot{x}_R \sin \omega t + \dot{y}_R \cos \omega t), \end{aligned} \tag{72c}$$

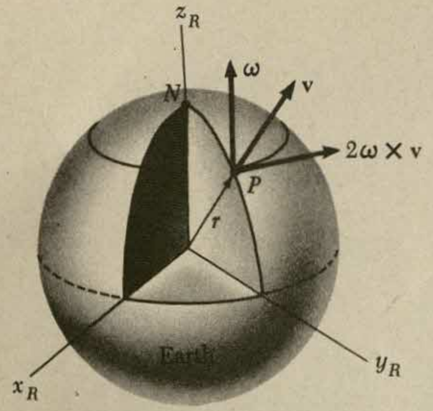
where we have used the fact that $\boldsymbol{\omega} = \omega_{z_I}\hat{\mathbf{z}}$ and we have projected $\mathbf{v}_R$ onto the y_I axis by following (65*b*). The last term in (72*a*) is

$$\begin{aligned} [\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}_R)]_{x_I} &= -\omega(\boldsymbol{\omega} \times \mathbf{r}_R)_{y_I} = -\omega^2(\mathbf{r}_R)_{x_I} \\ &= -\omega^2(x_R \cos \omega t - y_R \sin \omega t). \end{aligned} \tag{72d}$$

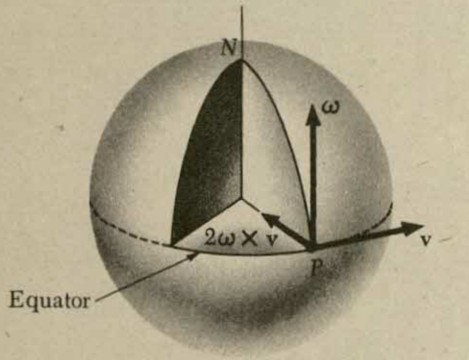
On adding up (72*b*), (72*c*), and (72*d*) we obtain

$$\begin{aligned} \ddot{x}_I &= (\mathbf{a}_R)_{x_I} - 2\omega(v_R)_{y_I} - \omega^2(\mathbf{r}_R)_{x_I} \\ &= \ddot{x}_R \cos \omega t - \ddot{y}_R \sin \omega t - 2\omega\dot{x}_R \sin \omega t - 2\omega\dot{y}_R \cos \omega t \\ &\quad - \omega^2 x_R \cos \omega t + \omega^2 y_R \sin \omega t, \end{aligned} \tag{73}$$

in agreement with (69*a*). This proves that the vector equation (72) is identical with the component equations (69).



Example of Coriolis acceleration in a rotating coordinate system: Rotating frame ($x_R y_R z_R$) is fixed on earth; ω is parallel to z_R. An object projected vertically upward from point P on surface of earth has initial velocity v. The Coriolis acceleration $2\omega \times v$ is tangent to the parallel of latitude through P, as shown; N is the North Pole. If object were allowed to fall freely from a height above the surface, the Coriolis acceleration would be oppositely directed. Why?



Another example of Coriolis acceleration: If object P at equator has velocity v (relative to the surface of the earth) tangent to equator, it will experience an acceleration toward the center of earth of magnitude $2\omega v$ in addition to acceleration of gravity g. What if v were directed to the west rather than to the east?

Equations (69) and (72*a*) are important. They relate the acceleration $\mathbf{a}_R$ observed in a rotating frame to the acceleration $\mathbf{a}_I$ in an inertial frame. We must have $\mathbf{a}_I$ in order to use Newton's second law. All quantities on the right-hand side of (72*a*) are as observed in the rotating frame. The first term $\mathbf{a}_R$ is the acceleration as observed in the rotating frame; the second term, $2\omega \times \mathbf{v}_R$, is called the *Coriolis acceleration;* and the third term, $\omega \times (\omega \times \mathbf{r}_R)$, is the usual centripetal acceleration of a particle in circular motion. We have omitted for simplicity a term $(d\omega/dt) \times \mathbf{r}_R$ which arises if the angular velocity ω is changing.

The fictitious force $\mathbf{F}_0$ for steady rotation ($\dot{\omega} = 0$) is given by (50) and (72*a*): From $F_0 = -M\mathbf{a}_0$ we have

$$\mathbf{F}_0 = \underset{\substack{\text{Coriolis}\\ \text{force}}}{-2M\omega \times \mathbf{v}_R} - \underset{\substack{\text{Centrifugal}\\ \text{force}}}{M\omega \times (\omega \times \mathbf{r}_R)}. \tag{74}$$

The centrifugal force is familiar; it is the only fictitious force acting on a particle which is at rest ($\mathbf{v}_R = 0$) in the rotating frame. We may rewrite the centrifugal force as

$$-M\omega \times (\omega \times \mathbf{r}_R) = M\omega^2 \boldsymbol{\rho}, \tag{74a}$$

where $\boldsymbol{\rho}$ is the vector from the axis to the particle, and normal to the axis. This is just our earlier result (55).

The Coriolis force is a fictitious force which acts on a particle only if it is in motion with respect to the rotating frame.

Example. Uniform linear motion, inertial frame. In an inertial frame a body moves freely with a trajectory given by

$$x_I = v_0 t; \qquad y_I = 0; \qquad z_I = 0. \tag{75}$$

What is the trajectory in a frame rotating with constant angular velocity ω counterclockwise about the z_I axis?

From (65) we have, with values from (75),

$$\begin{aligned} v_0 t &= x_R \cos \omega t - y_R \sin \omega t; \\ 0 &= x_R \sin \omega t + y_R \cos \omega t; \\ 0 &= z_R. \end{aligned} \tag{76}$$

We solve (76) to obtain

$$x_R = v_0 t \cos \omega t; \qquad y_R = -v_0 t \sin \omega t; \qquad z_R = 0. \tag{77}$$

Example. Free fall from tower. Consider a body falling under gravity after release from rest ($\mathbf{v}_R = 0$ at $t = 0$) at a point ($x_R{}^0$, 0, 0) near the earth's surface at the equator. The origin of the x_R, y_R, z_R system is at the center of the earth. The axis of rotation of the earth is the z axis. Calculate the coordinate y_R at which the body strikes the earth.

The true acceleration (due to gravity) is in the direction $-\hat{x}_R$ in the rotating frame (i.e., the frame fixed on the earth) and has the magnitude g. There is no true force component in the direction $\hat{y}_R$. The projection of the acceleration equation (72) on the direction $\hat{y}_R$ is therefore

$$0 = \ddot{y}_R + 2\omega\dot{x}_R - \omega^2 y_R, \tag{78}$$

and in lowest order we may set $\dot{x}_R = -gt$, where g is the acceleration of gravity, and neglect the term $-\omega^2 y_R$ because y_R is initially zero. Then

$$\ddot{y}_R \cong 2\omega gt; \qquad \dot{y}_R \cong \omega g t^2, \tag{79}$$

and the displacement is

$$y_R \cong \tfrac{1}{3}\omega g t^3. \tag{80}$$

This is to the east. For the rotation of the earth ω is positive.

This displacement is not difficult to observe. Let the release point be on a tower 100 m above the ground. Then, from the relation $\Delta x_R \cong -\frac{1}{2}gt^2$ for a freely falling body, we have

$$t^2 \cong -\frac{2\Delta x_R}{g} \approx \frac{2\times 10^4}{10^3} \approx 20\ \text{sec}^2, \tag{81}$$

whence, with $\omega \approx 0.7\times 10^{-4}\ \text{sec}^{-1}$ for the rotation of the earth,

$$y_R \cong \frac{1}{3}\omega g t^3 \approx \frac{(10^{-4})(10^3)(10^2)}{3} \approx 3\ \text{cm}. \tag{82}$$

How would you measure the true vertical from the tower, with respect to which the shift y is observed?

Example. Uniform linear motion, rotating frame. What force is needed to cause a body of mass M to move in a straight line at a uniform velocity through and normal to the axis, as viewed in a frame in uniform rotation?

The acceleration relation is

$$\mathbf{a}_I = \mathbf{a}_R + 2\boldsymbol{\omega}\times\mathbf{v}_R + \boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r}_R). \tag{83}$$

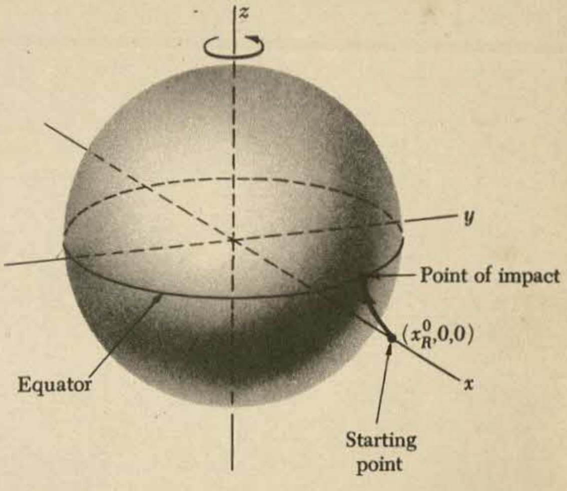
We require $\mathbf{a}_R = 0$ and $\mathbf{v}_R = v_0\hat{\boldsymbol{\rho}}_R$, where $\hat{\boldsymbol{\rho}}_R$ is a unit vector from the axis to the particle, and normal to the axis of rotation. Thus, if $\hat{\boldsymbol{\varphi}}_R$ is a unit vector in the circumferential direction,

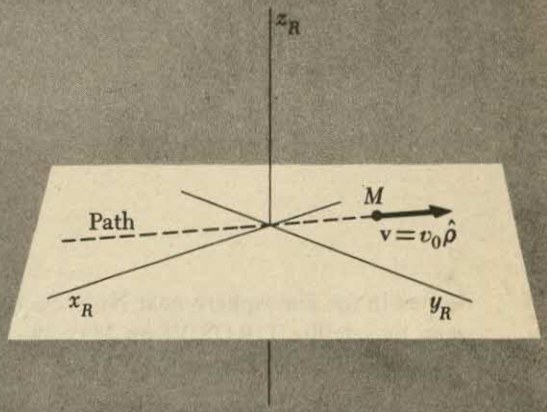
$$\mathbf{a}_I = 2\omega v_0\hat{\boldsymbol{\varphi}}_R - \omega^2 v_0 t\hat{\boldsymbol{\rho}}_R, \tag{84}$$

and

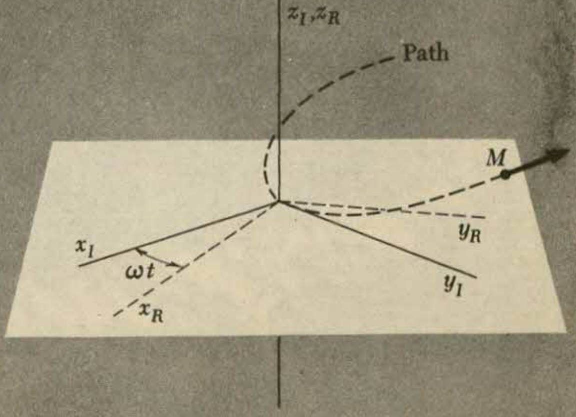
$$\mathbf{F} = 2M\omega v_0\hat{\boldsymbol{\varphi}}_R - M\omega^2 v_0 t\hat{\boldsymbol{\rho}}_R. \tag{85}$$

Thus we need a constant force in the circumferential direction $\hat{\boldsymbol{\varphi}}_R$ and the usual centripetal force inward in the radial direction $\hat{\boldsymbol{\rho}}_R$ normal to the axis.





Body of mass M moving in a straight line (in $x_r y_r$ plane) through axis of rotation, in frame which is uniformly rotating.



Same object viewed from inertial frame.

Example. Centrifugal correction to g. If g_R is the apparent acceleration of gravity at the equator in a frame rotating with the earth, what is the true value g_I of the acceleration of gravity as corrected for the centrifugal force?

We have for the effective acceleration in the rotating frame, with $\hat{\mathbf{z}}_R$ directed outward and normal to the surface of the earth,

$$\mathbf{F}_I + \mathbf{F}_0 = -Mg_R\hat{\mathbf{z}}_R; \tag{86}$$

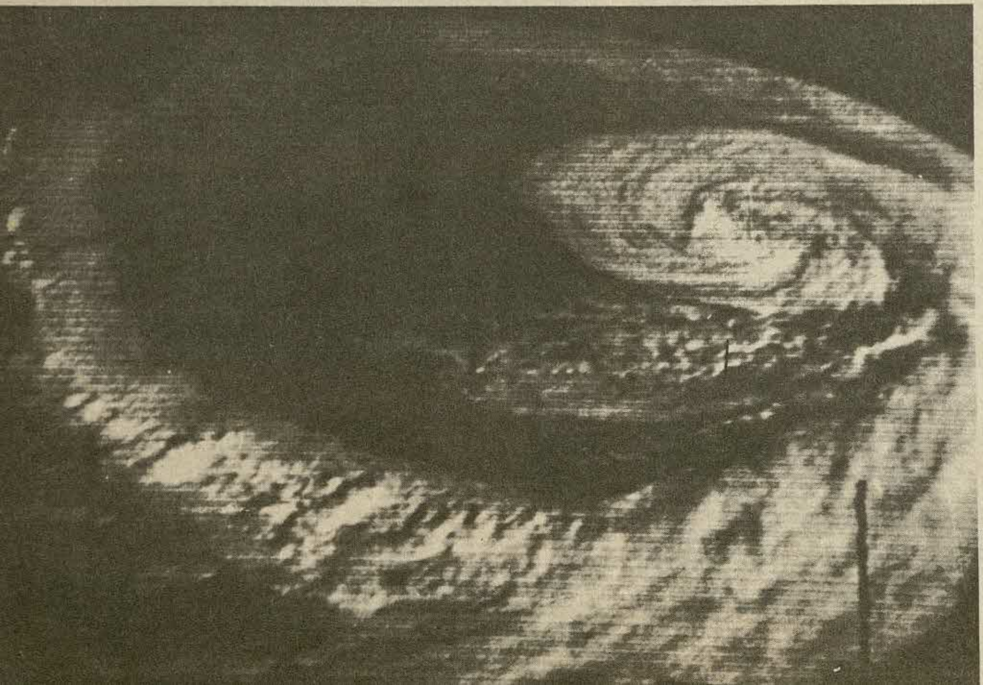
but $\mathbf{F}_0 = -M\mathbf{a}_0$, where $\mathbf{a}_0 = -\omega^2\boldsymbol{\rho}$, so that at an instant when $\hat{\mathbf{z}}_R$, $\hat{\mathbf{z}}_I$, and $\hat{\boldsymbol{\rho}}$ coincide,

$$(F_I)_{z_R} = -Mg_I = -M(g_R + \omega^2\rho). \tag{87}$$

At the equator $\rho = R$, so that

$$g_I = g_R + \omega^2 R. \tag{88}$$

Example. Direction of winds. Unequal heating of the earth's atmosphere in equatorial and polar regions produces a horizontal variation in pressure along north-south lines; yet winds have a dominant velocity component along east-west lines. This is explained by the rotation of the earth. Show that the steady motion of a nonviscous gas on the surface of the earth is parallel to isobars (lines of constant pressure). What is the wind pattern around a localized high-pressure zone (anticyclone) in the northern hemisphere?



Vortex in the atmosphere near Nova Scotia, as seen by satellite TIROS VI on May 29, 1963. (*Courtesy of Dr. F. Singer and NASA*)

The pressure gradient force acts north-south, perpendicular to the isobars. The rotation of the earth makes a north-south flow appear to have an east-west component, relative to the rotating earth. This problem can be understood directly from our earlier analysis of the Foucault pendulum.

In the northern hemisphere the wind blows clockwise around a high-pressure zone as the air moving radially outward from the high-pressure zone is deflected to the right as we look outward.

Reference: J. Spar, *Earth, sea, and air,* pp. 103–112 (Addison-Wesley, Reading, Mass., 1962.)

Mathematical Note 1. Differentiation of products of vectors

In Chap. 2 we considered the differentiation of vectors; recall in particular that if

$$\mathbf{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}}, \tag{89}$$

then

$$\dot{\mathbf{r}} = \dot{x}\hat{\mathbf{x}} + \dot{y}\hat{\mathbf{y}} + \dot{z}\hat{\mathbf{z}}, \tag{90}$$

provided that the basis vectors are constant in direction.

We now derive the relation

$$\frac{d}{dt}(\mathbf{A} \times \mathbf{B}) = \dot{\mathbf{A}} \times \mathbf{B} + \mathbf{A} \times \dot{\mathbf{B}}. \tag{91}$$

Let $\mathbf{P}(t)$ denote $\mathbf{A}(t) \times \mathbf{B}(t)$ and consider the expression

$$\begin{aligned}\mathbf{P}(t + \Delta t) - \mathbf{P}(t) &= \mathbf{A}(t + \Delta t) \times \mathbf{B}(t + \Delta t) - \mathbf{A}(t) \times \mathbf{B}(t)\\ &\cong \left[\mathbf{A}(t) + \frac{d\mathbf{A}}{dt}\Delta t\right] \times \left[\mathbf{B}(t) + \frac{d\mathbf{B}}{dt}\Delta t\right] - \mathbf{A}(t) \times \mathbf{B}(t)\\ &= \Delta t\left[\frac{d\mathbf{A}}{dt} \times \mathbf{B} + \mathbf{A} \times \frac{d\mathbf{B}}{dt}\right] + (\Delta t)^2\left[\frac{d\mathbf{A}}{dt} \times \frac{d\mathbf{B}}{dt}\right].\end{aligned} \tag{92}$$

Thus we have

$$\dot{\mathbf{P}} = \lim_{\Delta t \to 0} \frac{\mathbf{P}(t + \Delta t) - \mathbf{P}(t)}{\Delta t} = \dot{\mathbf{A}} \times \mathbf{B} + \mathbf{A} \times \dot{\mathbf{B}}. \tag{93}$$

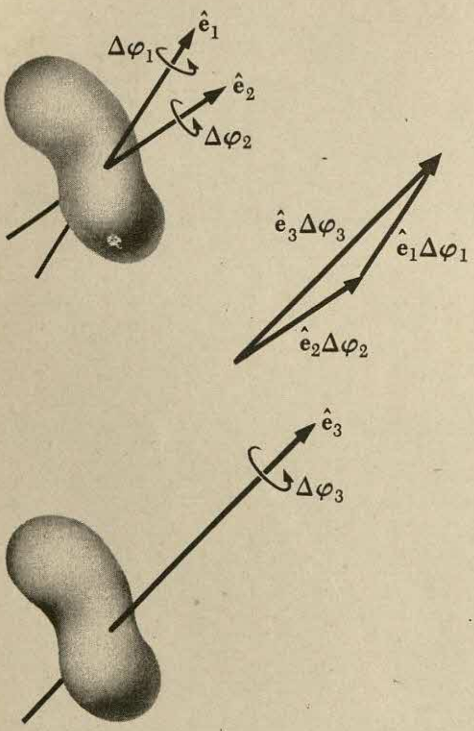
Note that the order of terms in the vector products is important in the result.

By a similar argument we have

$$\frac{d}{dt}(\mathbf{A} \cdot \mathbf{B}) = \dot{\mathbf{A}} \cdot \mathbf{B} + \mathbf{A} \cdot \dot{\mathbf{B}}. \tag{94}$$

Mathematical Note 2. Angular velocity as a vector quantity

In Chap. 2 we saw that finite rotations are not vectors, because two such rotations do not obey the vector law of addition. This difficulty does not arise in the limit of infinitesimal rotations, because the order in which two infinitesimal rotations is performed does not affect the final position of an object (except for terms that are as small as the square of the magnitude



of the infinitesimal rotations, and this is negligible in the limit). If a body is rotated through an infinitesimal angle $\Delta\varphi_1$ about an axis $\hat{e}_1$ and also about an axis $\hat{e}_2$ through an infinitesimal angle $\Delta\varphi_2$, the order in which the rotations take place does not affect the result, for sufficiently small $\Delta\varphi_1$ and $\Delta\varphi_2$. (We assume that all axes pass through a common point.) There exists a single rotation, about an axis $\hat{e}_3$ and through an angle $\Delta\varphi_3$, which is equivalent in the limit of infinitesimal $\Delta\varphi$ to the sum of the rotations 1 and 2. This rotation is given by the vector equation

$$\Delta\varphi_3\hat{e}_3 = \Delta\varphi_1\hat{e}_1 + \Delta\varphi_2\hat{e}_2. \tag{95}$$

If these infinitesimal rotations take place in an infinitesimal time Δt, we can define the vector angular velocity

$$\boldsymbol{\omega} \equiv \lim_{\Delta t\to 0} \frac{\Delta\varphi\hat{e}}{\Delta t}, \tag{96}$$

with $\boldsymbol{\omega}_3 = \boldsymbol{\omega}_2 + \boldsymbol{\omega}_1$. This defines a vector angular velocity which points along the instantaneous axis of rotation and whose magnitude is equal to the speed of rotation in radians per unit time.

To define the sense of the rotation (the direction of $\hat{e}$) we again adopt the convention of the right-hand rule: when the fingers of a right hand circle the axis of rotation in the direction of the rotation, then the thumb points in the direction of $\boldsymbol{\omega}$. The velocity of any point fixed in a rotating rigid body can be expressed simply in terms of the angular velocity $\boldsymbol{\omega}$. From a point 0 on the axis of rotation, a specific point on a rigid body is labeled by a vector $\mathbf{r}$ in the laboratory reference frame. A moment later, because of the rotation, this same point on the body is labeled by another $\mathbf{r}$, and the velocity $\mathbf{v} = \dot{\mathbf{r}}$ has a magnitude

$$v = \lim_{\Delta t\to 0}\frac{\Delta s}{\Delta t} = \rho\frac{d\varphi}{dt} = \rho|\boldsymbol{\omega}| = |\mathbf{r}|\sin\theta\,|\boldsymbol{\omega}|, \tag{97}$$

where θ is the polar angle between $\boldsymbol{\omega}$ and $\mathbf{r}$. The direction of $\mathbf{v}$ is perpendicular to both $\boldsymbol{\omega}$ and $\mathbf{r}$. All this information is comprised in the vector equation

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = \boldsymbol{\omega}\times\mathbf{r}, \tag{98}$$

for a point fixed in the rotating body.

Historical Note: The rotating bucket—Newton's view

The quotation which follows is a translation from Newton's *Principia Mathematica* published originally in 1686.

"The effects by which absolute and relative motions are distinguished from one another, are centrifugal forces, or those forces in circular motion which produce a tendency of recession from the axis. For in a circular motion which is purely relative no such forces exist; but in a true and absolute circular motion they do exist, and are greater or less according to the quantity of the [absolute] motion.

"For instance. If a bucket, suspended by a long cord, is so often turned about that finally the cord is strongly twisted, then is filled with water, and held at rest together with the water; and afterwards by the action of a second force, it is suddenly set whirling about the contrary way, and continues, while the cord is untwisting itself, for some time in this motion, the surface of the water will at first be level, just as it was before the vessel began to move; but, subsequently, the vessel, by gradually communicating its motion to the water, will make it begin sensibly to rotate, and the water will recede little by little from the middle and rise up at the sides of the vessel, its surface assuming a concave form. (This experiment I have made myself.)

". . . At first, when the *relative* motion of the water in the vessel was *greatest*, that motion produced no tendency whatever of recession from the axis; the water made no endeavor to move towards the circumference, by rising at the sides of the vessel, but remained level, and for that reason its *true* circular motion had not yet begun. But afterwards, when the relative motion of the water had decreased, the rising of the water at the sides of the vessel indicated an endeavor to recede from the axis; and this endeavor revealed the real circular motion of the water, continually increasing, till it had reached its greatest point, when *relatively* the water was at rest in the vessel . . .

"It is indeed a matter of great difficulty to discover and effectually to distinguish the *true* from the apparent motions of particular bodies; for the parts of that immovable space in which bodies actually move, do not come under the observation of our senses.

"Yet the case is not altogether desperate; for there exist to guide us certain marks, abstracted partly from the apparent motions, which are the differences of the true motions, and partly from the forces that are the causes and effects of the true motions. If, for instance, two globes, kept at a fixed distance from one another by means of a cord that connects them, be revolved about their common center of gravity, one might, from the simple tension of the cord, discover the tendency of the globes to recede from the axis of their motion, and on this basis the quantity of their circular motion might be computed. And if any equal forces should be simultaneously impressed on alternate faces of the globes to augment or diminish their circular motion, we might, from the increase or decrease of the tension of the cord, deduce the increment or decrement of their motion; and it might also be found thence on what faces forces would have to be impressed, in order that the motion of the globes should be most augmented; that is, their rear faces, or those which, in the circular motion, follow. But as soon as we knew which faces followed, and consequently which preceded, we should likewise know the direction of the motion. In this way we might find both the quantity and the direction of the circular motion, considered even in an immense vacuum, where there was nothing external or sensible with which the globes could be compared . . ."

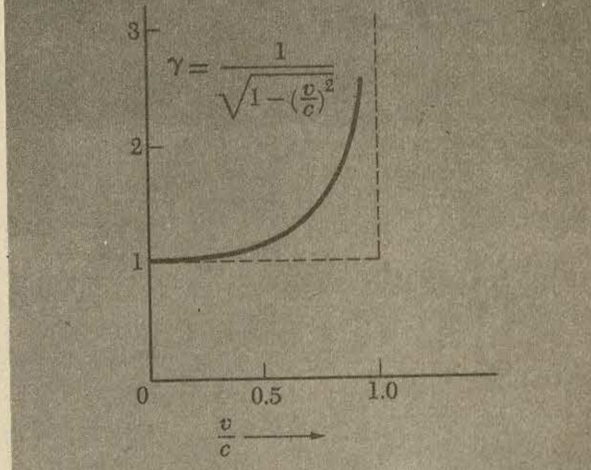
Chapter 4

Force on a charged particle; Gaussian units 93
Charged particle in uniform constant electric field 99
Examples: Longitudinal acceleration of proton 100
Longitudinal acceleration of electron 101
Transverse acceleration 101
Charged particle in uniform alternating electric field 102
Charged particle in constant magnetic field 104
Dimensions 106
Examples: Gyrofrequency 108
Gyroradius 108
180° magnetic focusing 109
Cyclotron acceleration principle 109
Further reading 110
Film list 110
Problems 111
Advanced Topic 1: Proton in crossed electric and magnetic fields 114
Example: Analogy to cycloid 115
Advanced Topic 2: Transformations of the reference frame 116
Mathematical Note: Complex numbers 118
Historical Note: Invention of the cyclotron 127

Simple Problems in Nonrelativistic Dynamics

This chapter briefly reviews three areas: the particle dynamics learned in high school physics, the elementary calculus learned in mathematics, and the vector analysis of Chap. 2. We set up and solve the equations of motion of several simple systems used in the laboratory. The systems that are discussed here relate to the motion of a charged particle in uniform electric and magnetic fields, an area extraordinarily rich in experimental applications. The chapter closes with detailed consideration of several useful transformations from one frame of reference to another.

We make the important assumption that the particle speed is always very much less than the speed of light; that is, we treat here only nonrelativistic problems and make only nonrelativistic transformations of reference frames. The effect of the special theory of relativity on our results will be treated in detail in Chaps. 12 and 13. Provided that $v \ll c$, our results are correct. We assume also that we are in a domain where the laws of classical physics are a sufficiently good approximation to the laws of quantum physics, which are the subject of Vol. IV.



Nonrelativistic mechanics gives an adequate description of the motion of a particle with speed v, if we can neglect the difference between $\frac{1}{\sqrt{1-v^2/c^2}}$ and 1. Here c is the velocity of light.

Force on a Charged Particle; Gaussian Units. The second law of Newton is, from Chap. 3,

$$\mathbf{F} = M\frac{d^2\mathbf{r}}{dt^2}, \tag{1}$$

if the mass M is constant. There are a number of important kinds of forces known in physics: the gravitational force, the electrostatic force, the magnetic force, and among others, various strong but short-range nuclear forces. Elementary mechanics can be a rather dull subject if the examples treated are limited to gravitational forces. Here and in the laboratory we want to consider problems involving electric and magnetic forces on a charged particle. A difficulty with our program is that a one-year high school physics course does not always treat electric and magnetic forces adequately. Somewhere at the back of the textbook these important forces may be described, but time may have run out before you came to them. If this happened to you, the best thing to do now is to go back and read through those parts of your high school book, or the PSSC book, which deal with electricity and magnetism. For comparison you should remember that the gravitational force on a particle of mass M at the surface of the earth is $\mathbf{F} = -Mg\hat{\mathbf{z}}$, where $\hat{\mathbf{z}}$ is the unit vector which points away from the center of the earth.

You will remember that like electrical charges repel each other with an inverse-square-law force directed along the line which connects the charges. This is Coulomb's law. The magnitude of the force is

$$F = \frac{q_1 q_2}{r^2}, \tag{2}$$

where q_1, q_2 are the charges. This is the form of Coulomb's law in the Gaussian system of units. Remember it in this form for use throughout the present course.

The dimensions of electric charge are determined by Eq. (2). The product q_1q_2 must have the same dimensions as r^2 times F, or [cm]^2 [dynes], which is [length]^2 [mass] [length]/[time]^2. Thus

$$[q]^2 = [\text{mass}]\,[\text{length}]^3/[\text{time}]^2,$$

so that the dimensions of charge are

$$[q] = [\text{mass}]^{1/2}\,[\text{length}]^{3/2}/[\text{time}].$$

We see that charge may be measured in gm$^{1/2}$ cm$^{3/2}$ sec^{-1} in CGS units. Two equal point charges which repel each other with a force of one dyne when they are one centimeter apart may be said to have each a charge of 1 gm$^{1/2}$ cm$^{3/2}$ sec^{-1}.

But this is a clumsy quantity to say and to write, and we greatly prefer to say instead that the value of this particular charge is 1 statcoulomb or 1 esu. That is, two equal point charges, each of one esu, repel each other with a force of one dyne when their separation is one cm. The statcoulomb or esu is the unit of charge in the Gaussian CGS system of units.† One also says that the statcoulomb is the unit of charge in electrostatic units (esu).

The charge q_p of a proton is denoted here and elsewhere in this course by the symbol e: $e = +4.8022 \times 10^{-10}$ esu. We call e the *elementary charge.* The charge q_e of an electron is equal to $-e$ by experiment.

From (2) the magnitude of the force between two protons at a separation of 10^{-12} cm is

$$F \cong \frac{(4.8 \times 10^{-10}\ \text{esu})^2}{(10^{-12}\ \text{cm})^2} \cong \frac{23 \times 10^{-20}}{10^{-24}} \cong 2.3 \times 10^5\ \text{dynes}. \qquad (3)$$

The force is repulsive between two protons. The magnitude of the force between a proton and an electron at the same separation of 10^{-12} cm is also given by (3), but the force is attractive because the charges are of opposite sign.

It is very convenient, especially for complex arrangements of many charges, to deal with the force per unit positive charge. To do this we imagine a positive test charge of magnitude 1 esu. We assume further that the test charge may be moved from point to point in space, leaving all other charges fixed. The force on the test charge is measured at any desired number of points; the measurement is to be made when the test charge is at rest at the point. Like any force, this force measured on a unit positive charge is a vector quantity; it is called the *electric field intensity* **E** at the point.

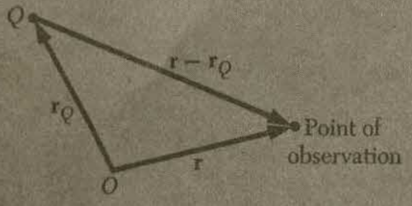
The units of **E** are those of force per unit charge. The force is in dynes and charge is in esu, so that the units of electric field intensity are dynes/esu. This is a perfectly good unit. It is common to use another name for it. We define an electric field of

$$1\,\frac{\text{statvolt}}{\text{cm}} \equiv 1\,\frac{\text{dyne}}{\text{esu}}.$$

The unit of statvolt/cm is introduced because it is convenient (as we shall see in detail in Chap. 5) to have a name (statvolt) for the product of electric field times distance. If the unit of the product is called the statvolt, then the unit of field will be called the statvolt/cm. We could get along without any special names for the units of charge and field, but it is very convenient to have them.

---

† Students who have not worked previously with Gaussian units should have no difficulty here: We have taken pains to introduce the units carefully in a liberal number of examples worked out in the text.



The electric field produced at position **r** by a charge Q at position $\mathbf{r}_Q$ is

$$\mathbf{E}(\mathbf{r}) = \frac{Q}{|\mathbf{r} - \mathbf{r}_Q|^3}(\mathbf{r} - \mathbf{r}_Q).$$

Just remember that a point charge 1 esu produces at a distance of 1 cm from the charge an electric field of 1 statvolt/cm.

We may construct a three-dimensional map of the electric field due to a static arrangement of charge. To each point of the map we associate a vector having the magnitude and direction of the electric field intensity **E**. It may be clearer to say that to each point of the map we associate a triplet of numbers, which are the values of the components E_x, E_y, E_z. Such a map is known as a vector field.

Suppose that we have mapped out the electric field **E** in a region of space. This is a useful quantity to know, for now we can write the force on any other charge of magnitude q in that region as

$$\mathbf{F} = q\mathbf{E}. \tag{4}$$

Here **F** will depend on the position of q because **E** is to be taken at that position. If **E** is produced by a single charge Q at the position $\mathbf{r}_Q$, then by (2) we can write the vector **E** at the point **r** as

$$\mathbf{E}(\mathbf{r}) = \frac{Q}{|\mathbf{r} - \mathbf{r}_Q|^3}(\mathbf{r} - \mathbf{r}_Q). \tag{5}$$

Here the cube of $|\mathbf{r} - \mathbf{r}_Q|$ appears in the denominator because in the numerator we have a vector of magnitude $|\mathbf{r} - \mathbf{r}_Q|$. If we take the origin at the charge Q, then $\mathbf{r}_Q = 0$ and we can rewrite (5) as

$$\mathbf{E}(\mathbf{r}) = \frac{Q}{r^3}\mathbf{r} = \frac{Q}{r^2}\hat{\mathbf{r}}. \tag{6}$$

Here the numerator on the right contains a unit vector $\hat{\mathbf{r}}$ from the origin pointing to the point **r** at which the field is measured. In terms of (6), the force on a charge q located at **r** is

$$\mathbf{F}(\mathbf{r}) = q\mathbf{E}(\mathbf{r}) = \frac{qQ}{r^2}\hat{\mathbf{r}}, \tag{7}$$

which is in agreement with (2).

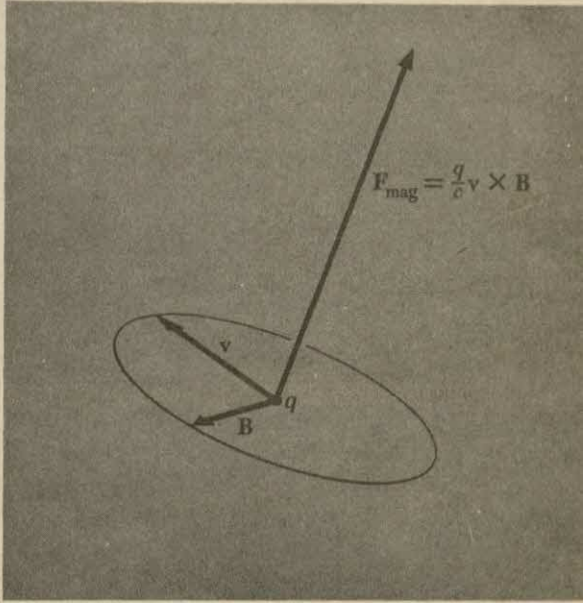
The electrostatic force

$$\mathbf{F}_{el} = q\mathbf{E} \tag{8}$$

is the contribution of the electric field to the force on a charge q. If the charge q is at rest relative to the observer, there may be no other force on the charge (omitting gravitational forces because they are usually so weak in comparison with electrostatic forces). But if the test charge or the charge q is moving, it is an experimental fact that there may be an additional contribution to the force on the charge. The additional force is found to be linearly proportional to the velocity v with which the charge moves relative to the observer, provided that the motion is at constant velocity. From experiment it is known that we can write the additional motional or magnetic force in the form

$$\mathbf{F}_{\text{mag}} = \frac{q}{c}\mathbf{v} \times \mathbf{B}, \tag{9}$$

where c is the speed of light, $\mathbf{v}$ is the particle velocity, and $\mathbf{B}$ is a vector quantity called the magnetic field intensity. We may if we wish consider **B** to be defined by this equation. This is discussed in full in Vol. II. Obviously the appearance of c in (9) is a matter of taste in the definition of **B** at this stage; one pleasant consequence of the form above is that **B** has the same dimensions as the electric field intensity **E**. That is, because we have divided v by c, we have the same dimensions for **B** as for **E**. The force $\frac{q}{c}\mathbf{v} \times \mathbf{B}$ on a charge in a magnetic field is just the force which causes a current-carrying wire to be pushed aside in a magnetic field perpendicular to the wire. There is a special name, the gauss, for the unit of magnetic field intensity in the Gaussian CGS system.


$\mathbf{F}_{\text{mag}} = \frac{q}{c}\mathbf{v} \times \mathbf{B}$
v
q
B

Consider the magnitude of the force on an electron moving in a magnetic field of 10,000 gauss, as might be produced by a small laboratory electromagnet. If the velocity of the electron is 3×10^8 cm/sec and is perpendicular to the magnetic field **B**, then the magnitude of the force is, according to (9),

$$F = (4.8 \times 10^{-10}\ \text{esu})\frac{(3 \times 10^8\ \text{cm/sec})}{(3 \times 10^{10}\ \text{cm/sec})}(1 \times 10^4\ \text{gauss})$$
$$= 4.8 \times 10^{-8}\ \text{dyne}.$$

It is obvious from this expression that the magnetic field has the dimensions of force per unit charge, just as the electric field does. But it is convenient to have a different name for the unit of magnetic field, and that is why we have the gauss for the magnetic field and the statvolt/cm for the electric field.

The source of the magnetic field may be a current loop or a solenoid or a permanent magnet. The direction of the magnetic force $\mathbf{F}_{\text{mag}}$ is normal to the plane formed by **v** and **B**. Later in this chapter we show that a charged particle moving in a magnetic field alone will travel in a circle (or, more generally, in a helix about an axis formed by the direction of the magnetic field). It is a simple matter to verify in the laboratory that a magnetic field applied perpendicular to the direction of the electron beam of the oscilloscope will deflect the beam in the direction mutually perpendicular to both **v** and **B**. The magnetic force, solenoids, and magnets are treated in detail in Vol. II.

The total force on a uniformly moving particle of charge q is the sum of the electrostatic and magnetic forces, (8) and (9). The sum, or (9) alone, is called the Lorentz force:

$$\boxed{\mathbf{F} = q\mathbf{E} + \frac{q}{c}\mathbf{v} \times \mathbf{B}.} \tag{10}$$

A great deal of physics comes out of Newton's second law $\mathbf{F} = M\mathbf{a}$ used together with Eq. (10); of course, an important part of the history of physics went into establishing these equations. Writing (10) down as an experimental fact here does not relieve us of the need to discuss it deeply in Vol. II.

In this chapter we assume that the kinetic energy of the particle is given by $\frac{1}{2}Mv^2$, and we treat only problems in which the electric and magnetic fields are uniform. (By *uniform* we mean independent of position; by *constant* we mean independent of time.) We defer the discussion of electrostatic potential and voltage until Chap. 5.

We shall need in this chapter the following numerical values: The speed of light:

$$c = 2.9979 \times 10^{10} \text{ cm/sec.}$$

The mass m of the electron:

$$m = 0.9107 \times 10^{-27} \text{ gm.}$$

The mass M_p of the proton:

$$M_p = 1.6724 \times 10^{-24} \text{ gm.}$$

In working with the Lorentz force (10) in Gaussian CGS units we express F in dynes; E in statvolts/cm; v in cm/sec; and B in gauss. The following conversion factors (derived in Vol. II) are used to convert the values of quantities expressed in MKS units to CGS units: We obtain F in dynes by multiplying by 10^5 the value of F in newtons. We obtain E in statvolts/cm by multiplying the value in volts/meter or newtons/coulomb by $10^6/c \cong \frac{1}{3} \times 10^{-4}$. If we are given E in volts/cm, we convert to statvolts/cm by multiplying the value in volts/cm by $10^8/c \cong 1/300$. We obtain B in gauss by multiplying by 10^4 the value in webers/m^2. Don't worry too much about these conversions at present; just use them when needed.

Charged Particle in Uniform Constant Electric Field. The equation for the force on a charge q and mass M in an electric field $\mathbf{E}$ uniform in space and constant in time is

$$\mathbf{F} = M\mathbf{a} = q\mathbf{E}. \tag{11}$$

From (11) we have

$$\mathbf{a} = \frac{d^2\mathbf{r}}{dt^2} = \frac{q}{M}\mathbf{E}, \tag{12}$$

an equation for the acceleration of the charge. This result is quite similar to that for the motion of a particle in a uniform gravitational field $\mathbf{F} = -Mg\hat{\mathbf{z}}$ at the surface of the earth, where $\hat{\mathbf{z}}$ is a unit vector which points out from the center of the earth. For the gravitational problem the equation of motion is $M\mathbf{a} = -Mg\hat{\mathbf{z}}$, or $\mathbf{a} = -g\hat{\mathbf{z}}$. If you have an initial difficulty in visualizing the motion of a particle in an electric field, think instead of the motion in a gravitational field pointing in the same direction. The latter motion is always treated exhaustively in high school physics. Equation (12) can be seen by trial or by direct integration to have the general solution

$$\mathbf{r}(t) = \frac{q\mathbf{E}}{2M}t^2 + \mathbf{v}_0t + \mathbf{r}_0. \tag{13}$$

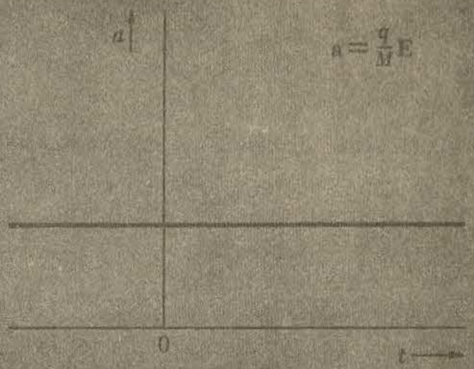
We recall the rule of differentiation that

$$\frac{d}{dt}t^n = nt^{n-1}. \tag{14}$$

You will naturally ask where the term $\mathbf{v}_0t + \mathbf{r}_0$, in the solution (13), comes from. It is clear that the expression

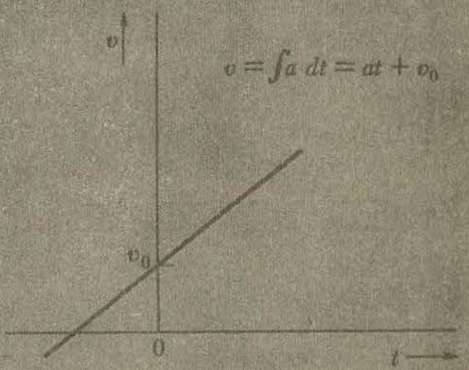
$$\mathbf{r}(t) = \frac{q\mathbf{E}}{2M}t^2 \tag{15}$$

will by itself satisfy Eq. (12). We call (12) a *differential equation* because it is an equation involving a derivative, in this instance the second derivative of $\mathbf{r}$ with respect to the time t. Equation (15) gives $\mathbf{r} = 0$ at $t = 0$. But there is nothing in the differential equation (12) which says that the particle must be at the origin at $t = 0$. The charge may have any initial position $\mathbf{r}_0$. We provide



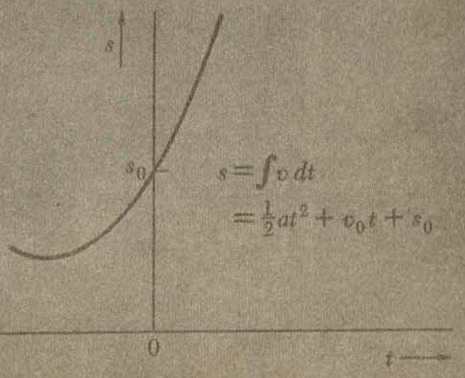
For a particle in a uniform field $\mathbf{E}$ the acceleration is constant:

$$\mathbf{a} = \frac{q}{M}\mathbf{E}.$$



the velocity increases linearly with time:

$$v = \int a\,dt = at + v_0.$$



and the displacement s is quadratic in the time:

$$s = \int v\,dt = \tfrac{1}{2}at^2 + v_0t + s_0.$$

for the assigned initial position in the solution for $\mathbf{r}$ by including a constant term $\mathbf{r}_0$:

$$\mathbf{r}(t) = \frac{q\mathbf{E}}{2M}t^2 + \mathbf{r}_0. \tag{16}$$

Because $\mathbf{r}_0$ is a constant, we can add it to the solution and the differential equation (12) is still satisfied.

The velocity $\mathbf{v}$ is obtained by differentiating $\mathbf{r}$ with respect to time. If we had (16) for $\mathbf{r}$, we would have

$$\mathbf{v}(t) = \frac{d\mathbf{r}}{dt} = \frac{d}{dt}\frac{q\mathbf{E}}{2M}t^2 = \frac{q\mathbf{E}}{M}t \tag{17}$$

for the velocity. This is zero at $t = 0$. But there is nothing in the differential equation (12) to compel the velocity to be zero at $t = 0$; the velocity can have any assigned initial value $\mathbf{v}_0$. We provide for the assigned initial velocity by adding a term $\mathbf{v}_0 t$ to the solution (16) for $\mathbf{r}$. This term does not disturb the differential equation nor does it change the initial value of position, because $\mathbf{v}_0 t = 0$ at $t = 0$. The term does satisfy the initial condition on the velocity, for now $\mathbf{r}$ is given by (13), whence

$$\mathbf{v}(t) = \frac{d\mathbf{r}}{dt} = \frac{d}{dt}\left(\frac{q\mathbf{E}}{2M}t^2 + \mathbf{v}_0 t + \mathbf{r}_0\right) = \frac{q\mathbf{E}}{M}t + \mathbf{v}_0. \tag{18}$$

We see that $\mathbf{v} = \mathbf{v}_0$ at $t = 0$, as required. The solution (13) satisfies the differential equation (12) and the *initial* or *boundary conditions* on the position and the velocity.

Example. Longitudinal Acceleration of Proton. A proton is accelerated from rest for 1 nanosecond ($= 10^{-9}$ sec) by an electric field $E_x = 1$ statvolt/cm. What is the final velocity?

The velocity is given by (18):

$$\frac{d\mathbf{r}}{dt} = \frac{e}{M}\mathbf{E}t + \mathbf{v}_0; \tag{19}$$

for our problem this† reduces to

$$v_x(t) = \frac{e}{M}E_x t; \qquad v_y = v_z = 0, \tag{20}$$

because we have specified that $\mathbf{v} = 0$ at $t = 0$. Thus the final

---

† Equation (19) is a vector equation; with $\mathbf{E} = (E_x, 0, 0)$ and $\mathbf{v}_0 = 0$ it reduces to the three component equations

$$\frac{dx}{dt} = \frac{e}{M}E_x; \qquad \frac{dy}{dt} = 0; \qquad \frac{dz}{dt} = 0.$$

velocity at $t = 1 \times 10^{-9}$ sec is, roughly,

$$v_x \approx \frac{(5 \times 10^{-10}\ \text{esu})(1\ \text{statvolt/cm})(1 \times 10^{-9}\ \text{sec})}{(2 \times 10^{-24}\ \text{gm})}$$

$$\approx 3 \times 10^5\ \text{cm/sec}. \quad (21)$$

Note that (1 esu) × (1 statvolt/cm) ≡ 1 dyne ≡ 1 gm-cm/sec². We have used 2×10^{-24} gm for the order of magnitude of the mass of the proton.

Example. Longitudinal Acceleration of Electron. An electron initially at rest is accelerated through 1 cm by an electric field of 1 statvolt/cm. What is the terminal speed?

From (18) we have, with $-e$ as the charge and m as the mass of the electron,

$$v_x(t) = -\frac{e}{m}E_xt; \qquad x(t) = -\frac{e}{2m}E_xt^2. \quad (22)$$

We want to eliminate t and solve for v_x in terms of x. It is convenient to form $v_x{}^2$ and then to rearrange factors using (22):

$$v_x{}^2 = \left(\frac{e}{m}E_xt\right)^2 = \left(\frac{2e}{m}E_x\right)\left(\frac{e}{2m}E_xt^2\right) = -\frac{2e}{m}E_xx$$

$$\approx \frac{2 \times 5 \times 10^{-10}}{10^{-27}}(1)(1) \approx (10^{-9})(10^{27})$$

$$\approx 10^{18}\ \text{cm}^2/\text{sec}^2. \quad (23)$$

Thus the final speed is

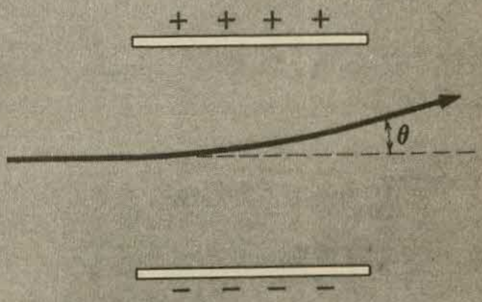
$$|v_x| \approx 10^9\ \text{cm/sec}. \quad (24)$$

The figure at the beginning of this chapter suggests that this speed may be considered to be nonrelativistic for many practical purposes (1 percent accuracy).

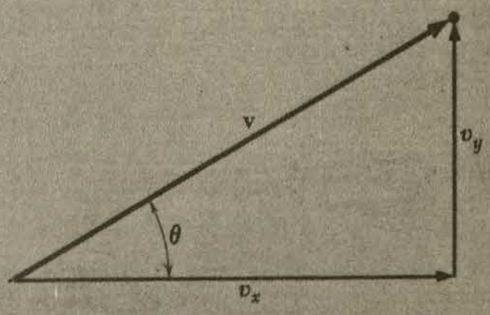
Example. Transverse Acceleration. After leaving the accelerating field E_x of the preceding example, the electron beam enters a region of length $L = 1$ cm in which there exists a transverse deflecting field $E_y = -0.1$ statvolt/cm. What angle with the x axis does the electron beam make on leaving the deflecting region?

Because there now is no x component of the field, the x component of the velocity will remain constant. The time τ spent by an electron in the deflecting region is given by

$$v_x\tau = L, \quad (25)$$



Deflection of electron beam by a transverse electric field.



$$\theta = \tan^{-1}\frac{v_y}{v_x}$$

Velocity vector $\mathbf{v} = \hat{x}v_x + \hat{y}v_y$.

or, if $v_x = 10^9$ cm/sec,

$$\tau = \frac{L}{v_x} = \frac{1}{10^9} = 10^{-9} \text{ sec.} \tag{26}$$

The transverse velocity acquired during this period is given by

$$v_y = -\frac{e}{m}E_y\tau \approx \frac{5 \times 10^{-10}}{10^{-27}}(10^{-1})(10^{-9}) \approx 5 \times 10^7 \text{ cm/sec.} \tag{27}$$

The angle θ which the terminal velocity vector makes with the x axis is given by $\tan\theta = v_y/v_x$, so that

$$\theta = \tan^{-1}\left(\frac{v_y}{v_x}\right) \approx \tan^{-1}\left(\frac{5 \times 10^7}{10^9}\right) = \tan^{-1} 0.05. \tag{28}$$

Now for a small angle we can make the approximation

$$\theta \cong \tan^{-1}\theta, \tag{29}$$

where θ is in radians. From (28) we see that $\theta \approx 0.05$ radian.

By estimating the value of the next term in the series expansion of $\tan^{-1}\theta$, we can check on the error in using the approximation (29). Standard mathematical tables give the series expansion of trigonometric functions. Thus Dwight† 505.1 reads

$$\tan^{-1} x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots, \text{ valid for } x^2 < 1. \tag{30}$$

The term $x^3/3$ for $x = 0.05$ is smaller than the leading term x by the factor $x^2/3 = (0.05)^2/3 \cong 10^{-3}$, or 0.1 percent. This error may be neglected if it is less than the experimental error in the measurement of θ. For small angles it is also true that $\sin\theta \cong \theta$ and $\cos\theta \cong 1 - \frac{1}{2}\theta^2$.

Charged Particle in Uniform Alternating Electric Field. Let

$$\mathbf{E} = \hat{\mathbf{x}}\, E_x = \hat{\mathbf{x}}\, E_x{}^0 \sin\omega t, \tag{31}$$

where $\omega = 2\pi f$ is the angular frequency and $E_x{}^0$ is the amplitude of the electric field vector. Often the superscript naught (^{0}) on the E is omitted if no ambiguity is introduced. The equation of motion is, from Eq. (12),

---

† For convenience nearly all references to mathematical tables will be to H. B. Dwight, *Tables of integrals and other mathematical data,* 4th ed. (Macmillan, New York, 1961). There are a number of excellent compact collections of mathematical data. You can get along well with any one of them. You should also have a handbook of chemical and physical data, such as the *Handbook of Chemistry and Physics,* Chemical Rubber Publishing Company, and a 10-in. slide rule.

$$\frac{d^2x}{dt^2} = \frac{q}{M}E_x = \frac{q}{M}E_x{}^0 \sin \omega t. \tag{32}$$

In solving differential equations we shall often use the excellent method of trial and error, guided by physical insight. We look for a solution of the form[†]

$$x(t) = x_1 \sin \omega t + v_0 t + x_0; \tag{33}$$

on differentiating (33) twice we find

$$\frac{d^2x}{dt^2} = -\omega^2 x_1 \sin \omega t. \tag{34}$$

The derivatives of the sine and cosine are given by

$$\frac{d}{d\theta}\sin\theta = \cos\theta; \qquad \frac{d^2}{d\theta^2}\sin\theta = -\sin\theta;$$

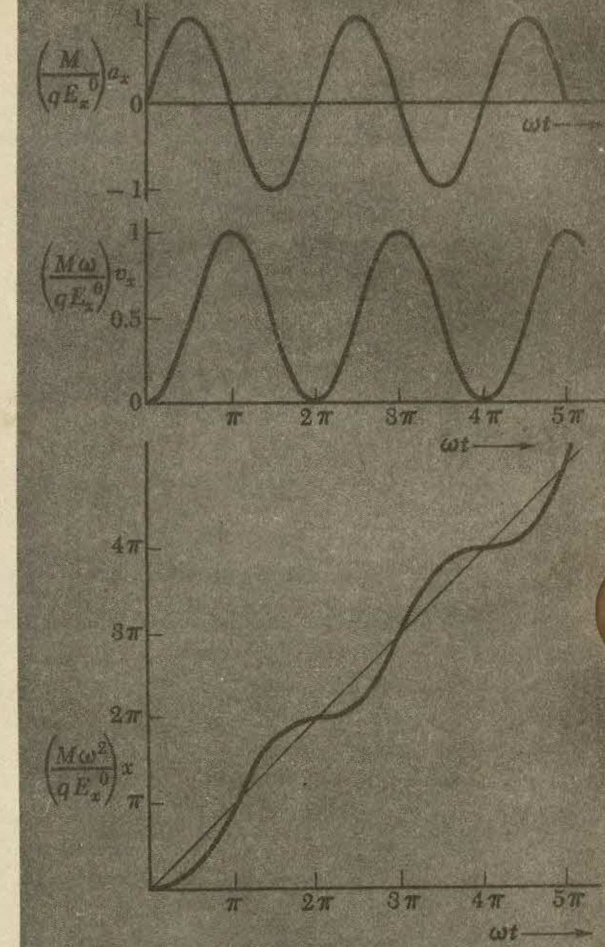
$$\frac{d}{d\theta}\cos\theta = -\sin\theta; \qquad \frac{d^2}{d\theta^2}\cos\theta = -\cos\theta.$$

Thus (33) is a solution of the equation of motion (32) provided that

$$-\omega^2 x_1 \sin \omega t = \frac{q}{M}E_x{}^0 \sin \omega t, \tag{35}$$

or

$$x_1 = -\frac{qE_x{}^0}{M\omega^2}. \tag{36}$$



For a charge q in field $\mathbf{E} = \hat{\mathbf{x}}E_x{}^0 \sin \omega t$, we have

$$a_x = \frac{q}{M}E_x{}^0 \sin \omega t$$

and

$$v_x(t) = \int a_x\,dt = -\frac{qE_x{}^0}{M\omega}\cos \omega t + v_0.$$

If $v_x(0) = 0$, then

$$v_0 = \frac{qE_x{}^0}{M\omega},$$

or

$$v_x(t) = \frac{qE_x{}^0}{M\omega}(1 - \cos \omega t), \ldots.$$

Now $x(t) = \int v_x(t)\;dt$

$$= \frac{qE_x{}^0}{M\omega}\int (1 - \cos \omega t)\;dt + x(0).$$

If $x(0) = 0$, then

$$x(t) = -\frac{qE_x{}^0}{M\omega^2}\sin \omega t + \frac{q}{M\omega}E_x{}^0 t.$$

[†] Here we participate in one of the common occupations of a physicist: finding the solution of a differential equation subject to prescribed initial conditions. This is an art in which intuitive guessing plays an important part. Often there are strictly prescribed mathematical procedures; commonly, the physicist asks himself "What could happen?" or "What else would you expect?" In the end the test is to substitute the guess in the original equation to see if the solution works. If the guess is wrong, try again. Intelligent guessing saves time, but even wrong guesses illuminate the problem.

Equation (32) states that the acceleration of a charged particle is a sinusoidal function of the time if the applied force is sinusoidal. Because the acceleration is oscillatory, the displacement at least in part must be oscillatory. For this reason we include in (33) such a term as sin ωt or cos ωt. We select sin ωt because two successive differentiations of a sine function give a sine function. The term x_0 must be included as the initial displacement. Since we must provide also for an initial velocity, we add a term $v_0 t$ which provides for any initial velocity, including zero. The effect of the term $v_0 t$ will persist at later times as a constant velocity superimposed on the oscillating one. The form $v_0 t$ is the only possibility; a higher power of t is not consistent with (32).

Note. We see from (38) that in this example, unlike the earlier examples, v_0 is not the only contribution to the velocity at $t = 0$.

On substituting (36) in (33), we have the result

$$x(t) = -\frac{qE_x{}^0}{M\omega^2}\sin\omega t + v_0 t + x_0. \tag{37}$$

The velocity is

$$v_x(t) = \frac{dx}{dt} = -\frac{qE_x{}^0}{M\omega}\cos\omega t + v_0; \tag{38}$$

thus at $t = 0$

$$v_x(0) = -\frac{qE_x{}^0}{M\omega} + v_0. \tag{39}$$

Do not confuse $v_x(0)$, which is the velocity at $t = 0$, with v_0, which is a constant to be selected to make $v_x(0)$ have the assigned value. If we choose the initial velocity to be zero we must have

$$v_0 = \frac{qE_x{}^0}{M\omega}; \tag{40}$$

on substitution of this result in (37) we have

$$x(t) = -\frac{qE_x{}^0}{M\omega^2}\sin\omega t + \frac{qE_x{}^0}{M\omega}t + x_0. \tag{41}$$

This is a somewhat unexpected result: with the boundary condition $v_x = 0$ at $t = 0$, the motion consists of an oscillation superposed on a constant drift velocity of $qE_x{}^0/M\omega$. This is because the particle never reverses its velocity for this special problem. The particle sidesteps continuously to the same side. Note that v_0 is *not* equal to $v_x(t = 0)$ in the present problem, but x_0 is equal to $x(t = 0)$.

Charged Particle in Constant Magnetic Field. The equation of motion of a charged particle of mass M and charge q in a constant magnetic field **B** is

$$M\frac{d^2\mathbf{r}}{dt^2} = M\frac{d\mathbf{v}}{dt} = \frac{q}{c}\mathbf{v} \times \mathbf{B}. \tag{42}$$

Let the magnetic field be directed along the z axis:

$$\mathbf{B} = \hat{\mathbf{z}}B. \tag{43}$$

Hence by the rule for the vector product

$$[\mathbf{v} \times \mathbf{B}]_x = v_y B; \qquad [\mathbf{v} \times \mathbf{B}]_y = -v_x B; \qquad [\mathbf{v} \times \mathbf{B}]_z = 0. \tag{44}$$

Thus (42) becomes

$$\dot{v}_x = \frac{q}{Mc} v_y B; \qquad \dot{v}_y = -\frac{q}{Mc} v_x B; \qquad \dot{v}_z = 0. \tag{45}$$

We see that the component of the velocity along the axis of the magnetic field, the z axis, is constant.

We can see directly another feature of the motion: The kinetic energy

$$K = \frac{1}{2} M v^2 = \frac{1}{2} M \mathbf{v} \cdot \mathbf{v} \tag{46}$$

is constant, for

$$\frac{dK}{dt} = \frac{1}{2} M(\dot{\mathbf{v}} \cdot \mathbf{v} + \mathbf{v} \cdot \dot{\mathbf{v}}) = M\mathbf{v} \cdot \dot{\mathbf{v}} = -M\mathbf{v} \cdot \left(\frac{q}{Mc} \mathbf{v} \times \mathbf{B}\right) \equiv 0, \tag{47}$$

because $\mathbf{v} \times \mathbf{B}$ is perpendicular to $\mathbf{v}$. Thus *a magnetic field does not change the kinetic energy of a free particle.*

Let us look for solutions† of the equations of motion (45) of the form

$$v_x(t) = v_1 \sin \omega t; \qquad v_y(t) = v_1 \cos \omega t. \tag{48}$$

This motion is circular. Now

$$\frac{dv_x}{dt} = \omega v_1 \cos \omega t; \qquad \frac{dv_y}{dt} = -\omega v_1 \sin \omega t, \tag{49}$$

so that (45) becomes

$$\omega v_1 \cos \omega t = \frac{qB}{Mc} v_1 \cos \omega t; \qquad -\omega v_1 \sin \omega t = -\frac{qB}{Mc} v_1 \sin \omega t. \tag{50}$$

These equations are satisfied if

$$\boxed{\omega = \frac{qB}{Mc} \equiv \omega_c;} \tag{51}$$

---

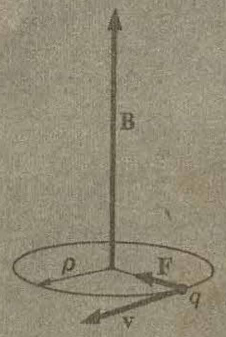
† Equation (47) tells us that K is a constant; we must conclude that $|\mathbf{v}|$ is constant also. This result suggests that we try a solution representing a uniform circular motion, in which x and y components of the velocity are sinusoidal with $\pi/2$ phase difference. It is convenient to represent qB/Mc as a single constant having dimensions of inverse time; the dimensions are easily seen from (45). We expect a solution involving a rotation for which this constant is related to the angular frequency ω.

this relation defines the *cyclotron frequency* (or *gyrofrequency*) ω_c as the frequency of the motion of the particle in the magnetic field Any value of v_1 will satisfy (50).

The cyclotron frequency can also be derived by an elementary argument. The inwardly directed magnetic force qBv_1/c provides the centripetal (inward) acceleration involved in the circular motion of the particle. The magnitude of the centrifugal acceleration is $v_1{}^2/r$ or $\omega_c{}^2r$, because $\omega_cr = v_1$. Thus

$$\frac{qBv_1}{c} = M\omega_c{}^2r, \tag{52}$$

whence $\omega_c = qB/Mc$.



Positive charge q, with initial velocity $\mathbf{v} \perp$ uniform **B**, describes a circle with constant speed v, radius $\rho = cMv/qB$.

What does the trajectory look like? The displacement is found by guessing or by integrating the expressions (48) with ω set equal to ω_c.

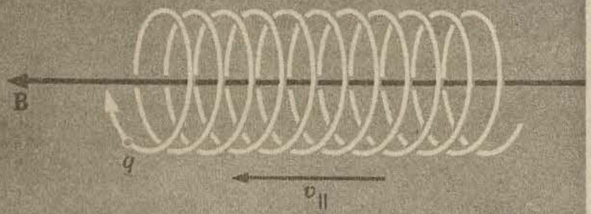
$$x(t) = x_0 - \rho \cos \omega_ct; \qquad x = x_0 - \frac{v_1}{\omega_c} \cos \omega_ct; \tag{53}$$

$$y(t) = y_0 + \rho \sin \omega_ct; \qquad y = y_0 + \frac{v_1}{\omega_c} \sin \omega_ct; \tag{54}$$

$$z(t) = z_0 + v_zt. \tag{55}$$

We see that the projection of the motion on the xy plane is a circle about the point x_0,y_0. The radius of the circle is

$$\rho = \frac{v_1}{\omega_c} = \frac{cMv_1}{qB}. \tag{56}$$



A positive charge q describes a helix of constant pitch in a uniform magnetic field **B**. The component of velocity $v_{\|}$ parallel to **B** is a constant.

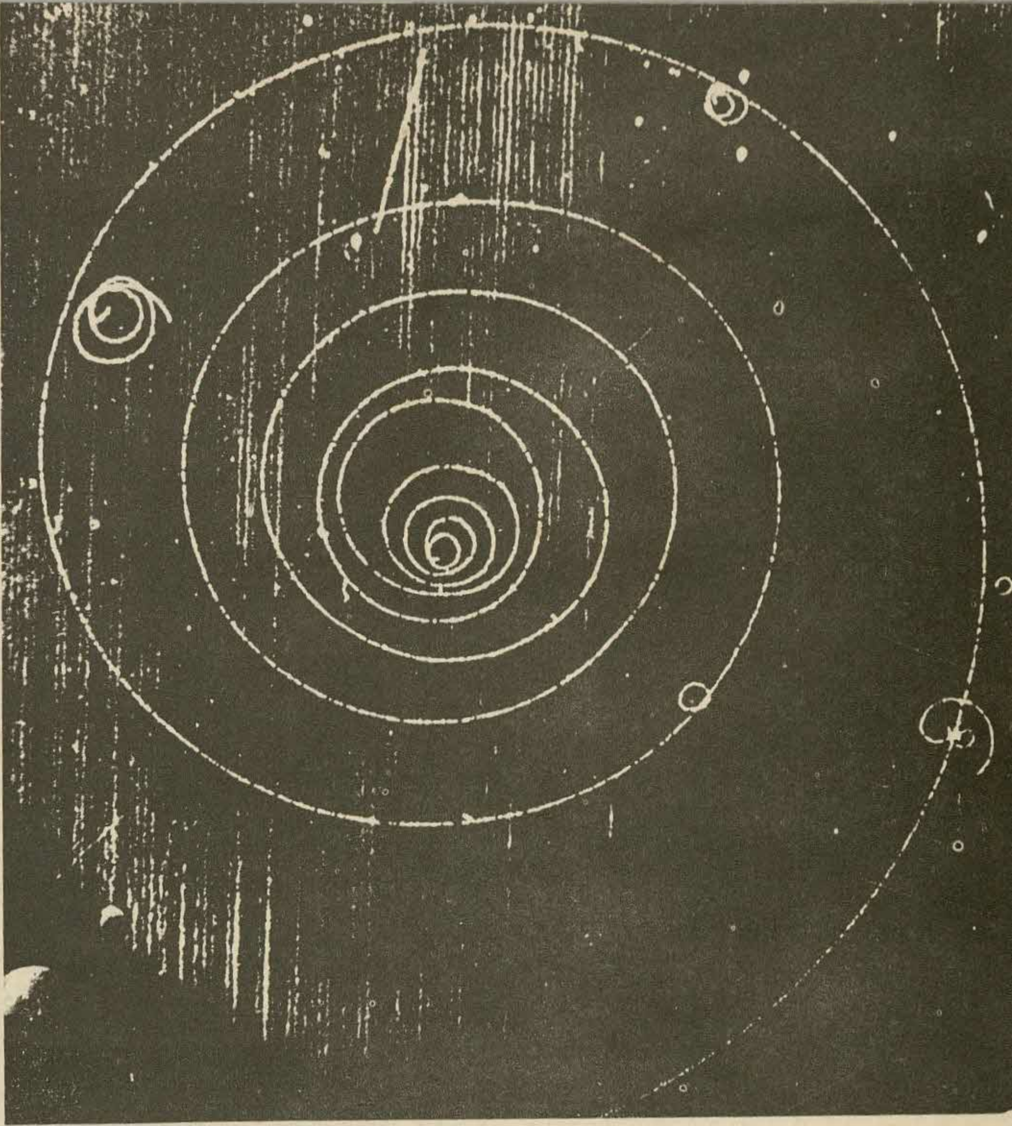
The radius ρ is sometimes called the *gyroradius* or *cyclotron radius.* The complete motion of the particle is a helix about the direction of **B** as axis; the velocity component parallel to **B** is constant.

Note that

$$\boxed{B\rho = \frac{cMv_1}{q},} \tag{57}$$

where Mv_1 is the momentum of the particle in the plane perpendicular to **B**. This is an important relation; we shall see in a later chapter that it is valid in the relativistic region if the relativistic momentum p is written for Mv_1. The relation may therefore be used to determine the momentum of a very high energy particle.

Dimensions. It is good practice always to check that the dimensions of both sides of a final equation are identical. This is an easy way to catch massive mistakes. On the right hand side of (57) we have



$$\left[\frac{cMv_1}{q}\right] = \left[\frac{L}{T}\right]\left[M\right]\left[\frac{L}{T}\right]\left[\frac{1}{q}\right] = \left[\frac{ML^2}{qT^2}\right], \tag{58}$$

where T denotes time and L denotes length. The brackets denote the dimensions of the quantity enclosed. On the left-hand side of (57) we have

$$[B\rho] = \left[\frac{F}{q}\right]\left[L\right] = \left[\frac{ML^2}{qT^2}\right], \tag{59}$$

because according to the Lorentz force equation (9) the dimensions of B in the Gaussian system of units are those of force divided by charge. We see that the dimensions in (58) are the same as in (59).

Hydrogen bubble chamber photograph of the path of a fast electron in a magnetic field. The electron enters at the lower left. The electron slows up by losing energy by ionization of hydrogen molecules. As the electron slows up its radius of curvature in the magnetic field decreases, hence the spiral orbit. (*Lawrence Radiation Laboratory*).

Example. Gyrofrequency. What is the gyrofrequency of an electron in a magnetic field of 10 kilogauss, or 1×10^4 gauss? (A field of 10 to 15 kilogauss is typical of ordinary laboratory iron core electromagnets.)

We have from (51)

$$\omega_c = \frac{eB}{mc} \approx \frac{(5 \times 10^{-10})(10^4)}{(10^{-27})(3 \times 10^{10})} \approx \frac{5 \times 10^{-6}}{3 \times 10^{-17}} \approx 2 \times 10^{11}\ \text{sec}^{-1}. \tag{60}$$

The corresponding frequency denoted by ν_c is

$$\nu_c = \frac{\omega_c}{2\pi} \approx 3 \times 10^{10}\ \text{cps}, \tag{61}$$

which is equivalent to an electromagnetic wavelength in free space of

$$\lambda_c = \frac{c}{\nu_c} \approx \frac{3 \times 10^{10}}{3 \times 10^{10}} \approx 1\ \text{cm}. \tag{62}$$

The gyrofrequency $\omega_c(p)$ of a proton is lower than that of an electron in the same magnetic field in the ratio 1/1836 of the electron mass to the proton mass. For a proton in a 10-kilogauss field:

$$\omega_c(p) = \frac{m}{M_p}\omega_c(e) \approx \frac{2 \times 10^{11}}{2 \times 10^3} \approx 10^8\ \text{sec}^{-1}. \tag{63}$$

The sense of rotation for the electron is opposite to that for the proton because their charges are opposite in sign.

Example. Gyroradius. What is the radius of the cyclotron orbit in a 10-kilogauss field for an electron of velocity 10^8 cm/sec normal to **B**?

We have for the gyroradius, using (60),

$$\rho = \frac{v_1}{\omega_c} \approx \frac{10^8}{2 \times 10^{11}} \approx 5 \times 10^{-4}\ \text{cm}. \tag{64}$$

The gyroradius for a proton of the same velocity is larger in the ratio M/m:

$$\rho \approx (5 \times 10^{-4})(2 \times 10^3) \approx 1\ \text{cm}. \tag{65}$$

180° Magnetic Focusing. Let a beam of charged particles enter a region in which there is a uniform magnetic field **B** perpendicular to the beam. The particles will be deflected with a radius of curvature given by the relation $B\rho = (c/q)Mv_t$, where v_t is the velocity component in the plane normal to **B**. If we examine the beam at some point, say after 180° of motion, we find it is spread out in the plane of the motion because the different particles with different masses and velocities have different radii of curvature.

The arrangement is used as a *momentum selector*, which is a device to obtain a beam of particles having closely equal momenta, if the particles all have the same charge q. One advantage of using a deflection of 180° is that particles of equal momenta, but moving through an entrance slit at slightly different angles, are brought to an approximately common focus after 180°.

The accuracy of the focusing is purely a problem in geometry. Consider a trajectory which makes initially an angle θ with the ideal trajectory. The distance from the entrance slit at which it will strike the target area is given by the chord C of the circle of radius ρ. The difference in length between the diameter and the chord is

$$2\rho - C = 2\rho(1 - \cos\theta) \approx \rho\theta^2, \tag{66}$$

where we have used for small θ the first two terms in the power series expansion of the cosine:

$$\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots, \tag{67}$$

as found in standard tables (Dwight 415.02). If we measure the angular focusing power by

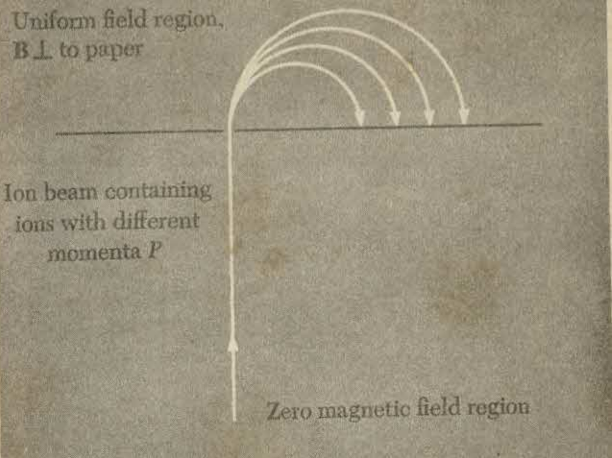
$$\frac{2\rho - C}{2\rho} \approx \frac{1}{2}\theta^2, \tag{68}$$

we have for $\theta = 0.1$ the value

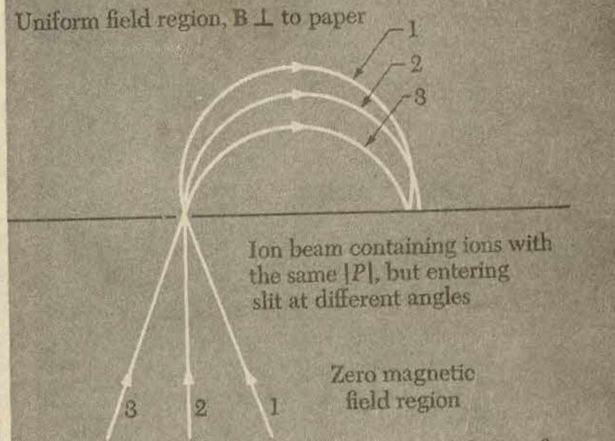
$$\frac{2\rho - C}{2\rho} \approx 5 \times 10^{-3}. \tag{69}$$

This exhibits the focusing action.

Cyclotron Acceleration Principle. Charged particles in a standard cyclotron move in roughly spiral orbits in a constant magnetic field as described in the historical note at the end of the chapter.



Magnetic field as a momentum selector.


Uniform field region, B ⊥ to paper
1
2
3
Ion beam containing ions with
the same |P|, but entering
slit at different angles
Zero magnetic
field region
3
2
1

180° focusing in a magnetic field. Ions of equal momenta but different directions are focused nearly together.

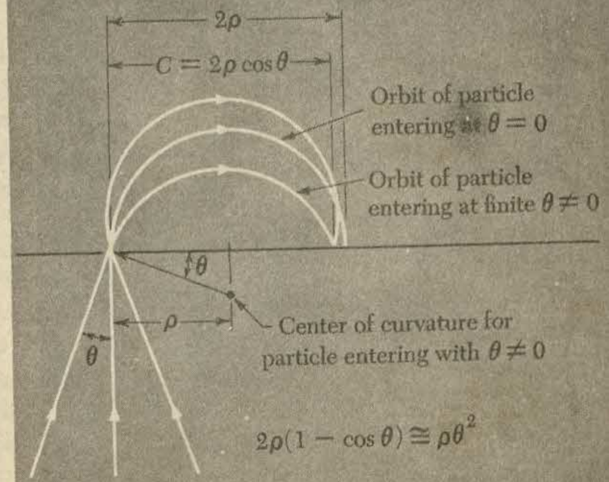
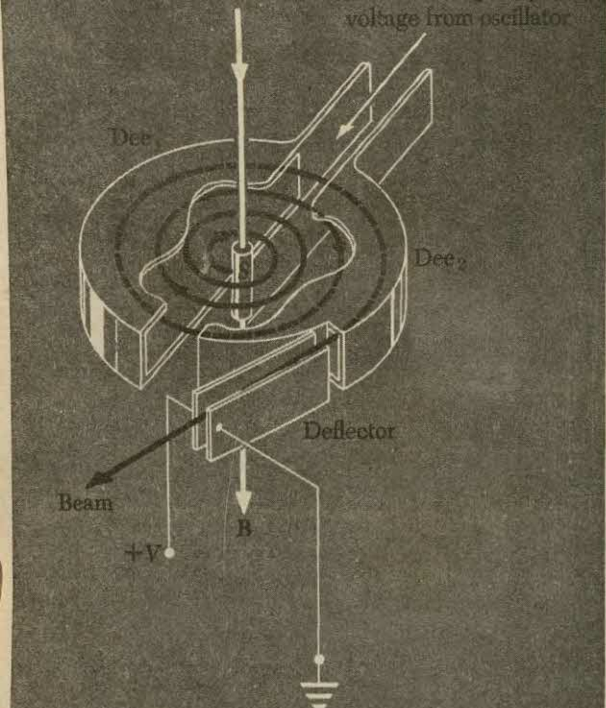


Diagram showing details of focusing in 180° velocity selector.



Cutaway view of a conventional low-energy cyclotron, consisting of ion source S, hollow accelerating electrodes (Dee$_1$, Dee$_2$), and deflector. The entire apparatus is immersed in a homogeneous vertical magnetic field **B** (pointed downward). The plane of the particle orbit is horizontal and is also the median plane of the dees. The accelerating rf electric field is confined to the gap between the dees.

The particles are accelerated every half-cycle (π radians) by an oscillating electric field. The requirement for periodic acceleration is that the frequency of the electric field should equal the cyclotron frequency of the particles.

The cyclotron frequency for protons in a magnetic field of 10 kilogauss is

$$\omega_c = \frac{eB}{M_p c} \approx \frac{(5 \times 10^{-10})(10^4)}{(2 \times 10^{-24})(3 \times 10^{10})} \approx 1 \times 10^8 \text{ sec}^{-1}, \quad (70)$$

or $f_c = \omega_c/2\pi \approx 10^7$ cps ≈ 10 Mc/sec. The frequency is independent of the energy of the particle so long as the velocity is nonrelativistic.

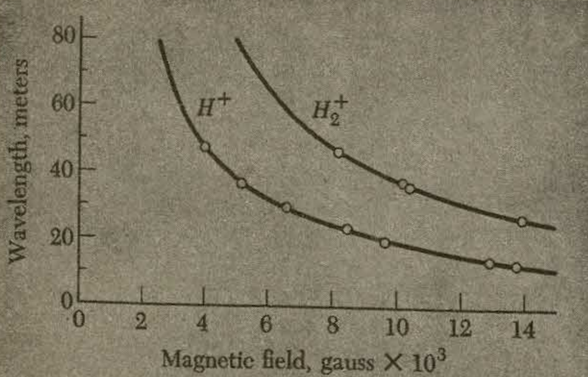
In each cycle of operation the particle picks up energy from the oscillating electric field. The effective radius of the orbit increases as the energy increases, for

$$r_c = \frac{v}{\omega_c} = \frac{\sqrt{2E/M_p}}{\omega_c}, \quad (71)$$

where E now denotes the energy. The energy of a nonrelativistic proton in a constant magnetic field is set by the outer radius of the cyclotron: at $\omega_c = 1 \times 10^8$ sec^{-1} and $r_c = 50$ cm, we have $v = \omega_c r_c \approx 5 \times 10^9$ cm/sec, or

$$E = \tfrac{1}{2} M_p v^2 \approx 10^{-24}(5 \times 10^9)^2 \approx 25 \times 10^{-6} \text{ ergs}. \quad (72)$$

This velocity is in practice sufficiently nonrelativistic for the operation of a conventional cyclotron.



Resonance condition in the first cyclotron (11-in. diameter). On the vertical scale is plotted the wavelength in free space of the rf power supplied to the accelerating electrodes (dees). The curves are the theoretical relations for H$^+$ and H$_2{}^+$ ions; the circles are the experimental observations. (Lawrence and Livingston, *Phys. Rev.* 40, 19, 1932.)

Further reading

PSSC, chaps. 24 (paragraphs 1–5), 28 (paragraphs 3–5), 29 (paragraphs 1, 2, 7, 12), and 30 (paragraphs 2, 3, 8, 9).

L. Hopf, *Introduction to the differential equations of physics*, trans. by W. Nef (Dover, New York, 1948). A compact and pleasant introduction to differential equations that requires little mathematical preparation and is well suited for independent study.

Film List

"Mass of the Electron" (18 min) E. Rogers (PSSC-MLA 0413). Shows the deflection of an electron.

Problems

Note: Always give the units with your answers; an answer without units is no answer at all.

1. *Proton in electric field—force.* What is the value in dynes of the force on a proton of charge e in an electric field $E = 100$ statvolts/cm? *Ans.* 4.8×10^{-8} dynes.

2. *Proton in magnetic field—force.* Under the following conditions what is the value (in dynes) and direction of the force on a proton of charge e in a magnetic field $\mathbf{B}$ of magnitude 100 gauss directed along the z axis:

(*a*) When the proton is at rest? *Ans.* 0 dynes.

(*b*) When the proton moves along the x axis with velocity $\mathbf{v} = 10^8\,\hat{\mathbf{x}}$ cm/sec? *Ans.* $-1.6 \times 10^{-10}\,\hat{\mathbf{y}}$ dynes.

3. *Kinetic energy of electron.* (*a*) What are the values of the momentum (in gm-cm/sec) and of the kinetic energy (in ergs) of an electron with velocity $\mathbf{v} = 10^8\,\hat{\mathbf{x}}$ cm/sec? (Recall that the kinetic energy of a particle of mass M is given by $\frac{1}{2}Mv^2$.)

(*b*) An electron is accelerated from rest by an electric field of 0.01 statvolts/cm acting for 5 cm. What is its final kinetic energy, in ergs? *Ans.* 2.4×10^{-11} ergs.

4. *Free motion of electron.* Given that at $t = 0$ the velocity of an electron is $\mathbf{v} = 10^6\,\hat{\mathbf{x}}$ cm/sec, and the position is $\mathbf{r} = 100\,\hat{\mathbf{y}}$ cm; find the position vector at $t = 0.1$ sec. No external fields are present. *Ans.* $10^5\,\hat{\mathbf{x}} + 10^2\,\hat{\mathbf{y}}$ cm.

5. *Electron in electric field—acceleration.* Given the data of Prob. 4 at $t = 0$, but now with an electric field $\mathbf{E} = 10^{-2}\,\hat{\mathbf{x}}$ statvolts/cm; find the position and velocity vectors at $t = 1 \times 10^{-8}$ sec.

6. *Electron in magnetic field—acceleration.* Given the data of Prob. 4 at $t = 0$, but now with $\mathbf{B} = 100\,\hat{\mathbf{z}}$ gauss and $\mathbf{E} = 0$:

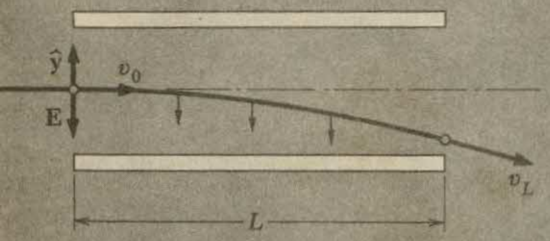
(*a*) Solve for the position vector at $t = 1 \times 10^{-8}$ sec.

(*b*) Find the momentum vector at both times $t = 0$ and $t = 1 \times 10^{-8}$ sec; give its magnitude and direction.

7. *Electric field to counterbalance gravity.* What value of the electric field intensity will give an electron an acceleration equal to 980 cm/sec², the acceleration of gravity? *Ans.* 1.86×10^{-15} statvolts/cm.

8. *Ratio of electric and gravitational forces between two electrons.* The magnitude of the electrostatic force between two electrons is e^2/r^2; the magnitude of the gravitational force is Gm^2/r^2, where $G = 6.67 \times 10^{-8}$ dyne-cm²/gm². What is the order of magnitude of the ratio of the electrostatic and gravitational forces between two electrons? *Ans.* 10^{42}.

9. *Crossed electric and magnetic fields* (see Advanced Topic 1). A charged particle moves in the x direction through a region in which there is an electric field E_y and a perpendicular magnetic field B_z. What is the condition necessary to ensure that the net force on the particle will be zero? Show the $\mathbf{v}$, $\mathbf{E}$, and $\mathbf{B}$ vectors on a diagram. What is the condition on v_x if $E_y = 10$ statvolts/cm and $B_z = 300$ gauss? *Ans.* $v_x = 1 \times 10^9$ cm/sec.



10. *Deflection between condenser plates.* A particle of charge q and mass M with an initial velocity $v_0\hat{x}$ enters an electric field $-E\hat{y}$. We assume **E** is uniform, i.e., its value is constant at all points in the region between plates of length L (except for small variations near the edges of the plates which we shall neglect).

(*a*) What forces act in the x and y directions, respectively?

Ans. $F_x = 0$; $F_y = -qE\hat{y}$.

(*b*) Will a force in the y direction influence the x component of the velocity?

(*c*) Solve for v_x and v_y as functions of time, and write the complete vector equation for $\mathbf{v}(t)$. *Ans.* $v_0\hat{x} - \frac{qE}{M}t\hat{y}$.

(*d*) Choose the origin at the point of entry, and write the complete vector equation for the position of the particle as a function of time while the particle is between the plates.

11. *Continuation of preceding problem.* If the particle in Prob. 10 is an electron of initial energy 10^{-10} ergs, if the electric field strength is 0.01 statvolts/cm, and if $L = 2$ cm, find:

(*a*) The vector velocity as it leaves the region between the plates.

(*b*) The angle $(\mathbf{v},\hat{x})$ for the particle as it leaves the plates. *Ans.* 2.7°.

(*c*) The point of intersection between the x axis and the direction of the particle as it leaves the field. *Ans.* 1.0 cm.

12. *Electric field of point charge.* The electric field intensity E at a distance r from a point charge q is given in Gaussian units by q/r^2 and is directed radially outward for q positive and inward for q negative.

(*a*) Write a vector relation for **E**.

(*b*) A charge $q = 2e$ corresponding to the charge on two protons is placed at the point $A(1,2,3)$.[†] Find the electric field intensity at the point $B(4,5,10)$. The unit of length is the centimeter.

13. *Transit time of ions.* A pulse of single charged cesium ions Cs^+ is accelerated from rest by an electric field of 1 statvolt/cm acting for 0.33 cm and afterward travels 1 mm in 87×10^{-9} sec in an evacuated field-free space.

(*a*) Derive from these data a value of the atomic mass of Cs^+.

Ans. 2.4×10^{-22} gm.

Compare with values you will find in tables, handbooks, or chemistry textbooks.

(*b*) What would be the time for protons to transit the 1-mm region?

Ans. 7.2×10^{-9} sec.

(*c*) For deuterons? $M_d \cong 3.2 \times 10^{-24}$ gm. A deuteron is a proton plus a neutron.

(*d*) Could one distinguish between the deuteron and the alpha particle in the experiment? The charge on the α particle is $2e$.

[†] The notation $A(1,2,3)$ refers to a point A whose Cartesian coordinates are 1, 2, and 3 respectively.

14. *Charged particles in uniform magnetic field.* An electron and a proton are accelerated by an electric field of 1 statvolt/cm acting for 10 cm; they then enter a uniform magnetic field of 10,000 gauss in a plane perpendicular to the field.

(*a*) What is the cyclotron frequency of each particle?

(*b*) What is the radius of the circular orbit of each particle?

Ans. $R_e = 1.8 \times 10^{-2}$ cm; $R_p = 0.77$ cm.

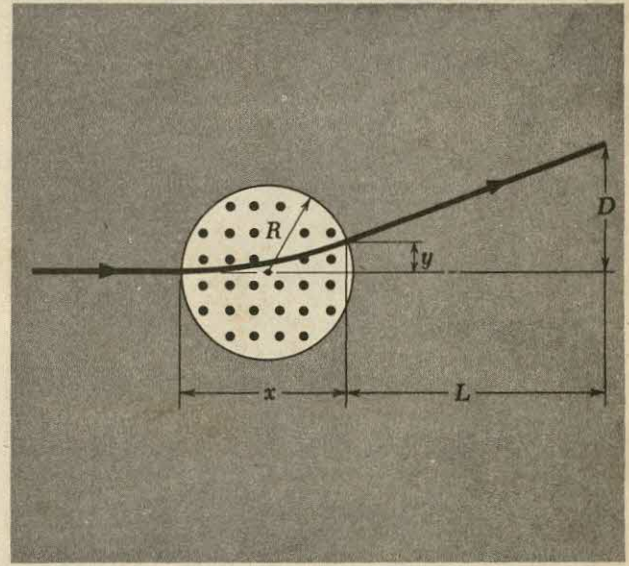
15. *Magnetic monopoles.* Magnetic monopoles or free poles have never (to the date this is written) been observed experimentally. An argument has been given by Dirac that the strength of a magnetic pole (denoted usually by g) if it exists should be an integral multiple of $\frac{1}{2} \cdot \frac{\hbar c}{e^2}$, or $\frac{137}{2}$, times the elementary electric charge e. If we take the integer to be unity, then

$$g = \frac{1}{2} \cdot \frac{\hbar c}{e^2} \cdot e \cong 3.3 \times 10^{-8} \text{ esu.}$$

If magnetic monopoles can be produced in collisions of particles at high energies, we believe that equal numbers of north and south monopoles will be produced.

(*a*) What is the value in dynes of the force $\mathbf{F} = g\mathbf{B}$ on a Dirac monopole at rest in a magnetic field $B = 5 \times 10^4$ gauss? *Ans.* 1.6×10^{-3} dyne.

(*b*) Compare with the electric force $\mathbf{F} = e\mathbf{E}$ on a proton at rest in an electric field of 5×10^4 volts/cm. Remember to convert E to statvolts/cm ($E \cong 5 \times 10^4/300$ statvolts/cm). *Ans.* 8.0×10^{-8} dyne.



16. *Magnetic deflection of electron beam.* Deflection of an electron beam in a cathode-ray tube may be accomplished by magnetic as well as by electrostatic means. A beam of electrons of energy W enters a region of transverse uniform magnetic field of strength B. (Neglect fringe effects.)

(*a*) If x is the distance from the point at which the electron entered the field to the point at which it leaves it, show that

$$y = r\left[1 - \sqrt{1 - \left(\frac{x}{r}\right)^2}\right],$$

where r is the radius of curvature of the electron in the transverse magnetic field. The radius of curvature is the radius of the circle which will match (coincide with) the curved portion of the path.

(*b*) If R is the radius of the magnet poles, then $x \cong 2R$ when $r \gg R$. Use the binomial expansion to show $y \approx 2R^2/r$.

17. *Acceleration in a cyclotron.* Suppose in a cyclotron that $\mathbf{B} = \hat{\mathbf{z}}B$ and

$$E_x = E \cos \omega_c t; \qquad E_y = -E \sin \omega_c t; \qquad E_z = 0,$$

with E constant. (In an actual cyclotron the electric field is not uniform in space.) We see that the electric field intensity vector sweeps around a circle with angular frequency ω_c. Show that the displacement of a particle is described by

$$x(t) = \frac{qE}{M\omega_c^2}(\omega_c t \sin \omega_c t + \cos \omega_c t - 1);$$

$$y(t) = -\frac{qE}{M\omega_c^2}(\omega_c t \cos \omega_c t - \sin \omega_c t),$$

where at $t = 0$ the particle is at rest at the origin. Sketch the first few cycles of the displacement.

Advanced Topic 1. Proton in crossed electric and magnetic fields

This is an important example which can be solved rather easily. Let $\mathbf{B} = \hat{\mathbf{z}}B$; if $\mathbf{E}$ is perpendicular to $\mathbf{B}$ we have, from the Lorentz force equation (10) and the definition (51) of ω_c, the equations of motion

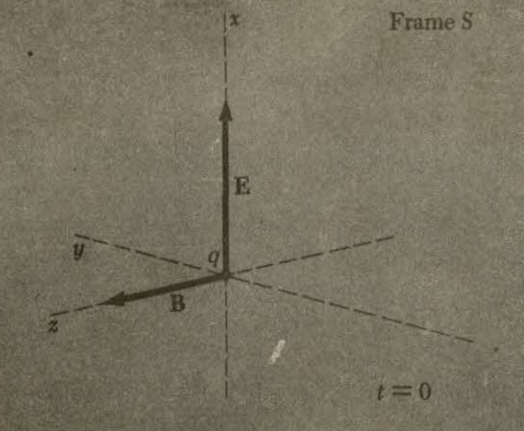
$$\dot{v}_x = \frac{e}{M_p}E_x + \omega_c v_y; \qquad \dot{v}_y = \frac{e}{M_p}E_y - \omega_c v_x; \qquad \dot{v}_z = 0. \qquad (73)$$

The electric field is in the xy plane and the magnetic field is parallel to the z axis. A *special solution* of these equations is, with $\omega_c = eB/M_pc$ and $\mathbf{E} = \hat{\mathbf{x}}\,E$,

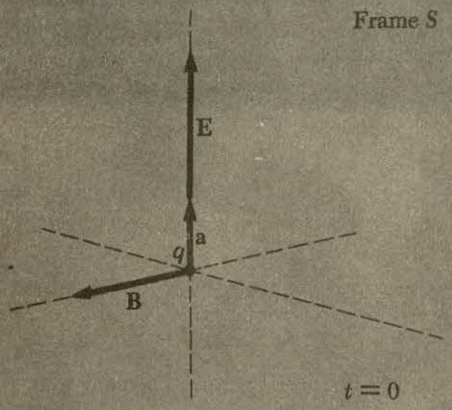
$$v_y = -\frac{cE}{B}; \qquad v_x = 0; \qquad v_z = \text{const}. \qquad (74)$$

We found this solution by curiosity, by looking for a solution with $\dot{\mathbf{v}} \equiv 0$. Note that in this special case the particle does not accelerate! The crossed electric and magnetic fields here act as a *velocity selector* for particles of velocity v_y given by (74); such particles pass undeflected through the apparatus. This solution suggests that we look for a general solution of (73) of the form

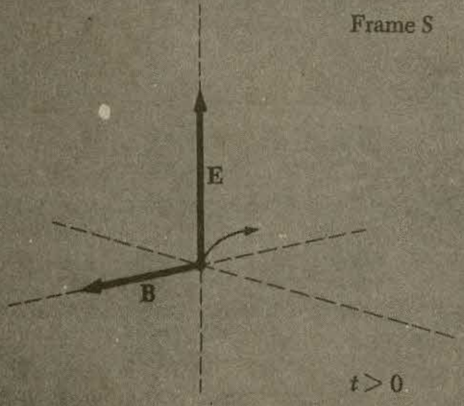
$$v_x(t) = \eta_x(t); \qquad v_y(t) = \eta_y(t) - \frac{cE}{B}. \qquad (75)$$



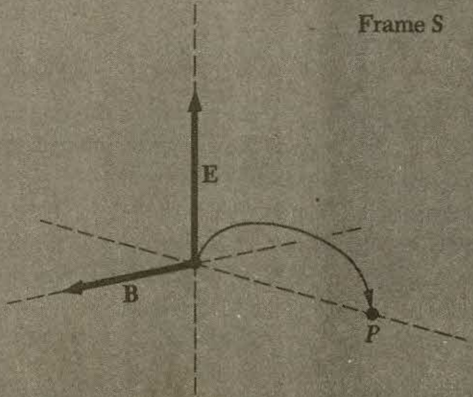
Consider positive charge q at rest at the origin in crossed **E** and **B** fields.



The initial acceleration of q is $\mathbf{a} = \dfrac{q\mathbf{E}}{M}$.



As soon as q acquires velocity in the direction of **E**, it experiences a force $\mathbf{F} = (q/c)\mathbf{v} \times \mathbf{B}$. The orbit then curves in the $(-y)$ direction . . .



and q ultimately comes to rest at P, a point on the y axis. It then begins a new cycle of motion.

Here η is the Greek character eta. On substituting (75) into (73), we have

$$\dot{\eta}_x = \omega_c\eta_y; \qquad \dot{\eta}_y = -\omega_c\eta_x. \tag{76}$$

This set of equations is identical with (45). Therefore

$$v_x(t) = v_1 \sin \omega_c t; \tag{77a}$$

$$v_y(t) = v_1 \cos \omega_c t - \frac{cE}{B}; \tag{77b}$$

$$v_z(t) = \text{const.} \tag{77c}$$

The motion described by (77) will appear circular (or helical, if $v_z \neq 0$) to an observer moving with the constant velocity

$$\mathbf{V} = \left(0, -\frac{cE}{B}, 0\right) \tag{78}$$

in the y direction. From this moving frame the proton appears to see only a magnetic field and zero electric field. Compare (78) with (74).

Example. Analogy to Cycloid. The motion in the laboratory reference system is similar to that of a point on the circumference of a wheel rolling in the y direction without slipping (i.e., a common cycloid) if the distance

$$V\frac{2\pi}{\omega_c} = \left(\frac{cE}{B}\right)\left(\frac{2\pi Mc}{qB}\right) \tag{79}$$

advanced in a period $(2\pi/\omega_c)$ is equal to the circumference $2\pi\rho$. By (57) we have for the circumference

$$2\pi\rho = 2\pi\frac{cMv_1}{qB}. \tag{80}$$

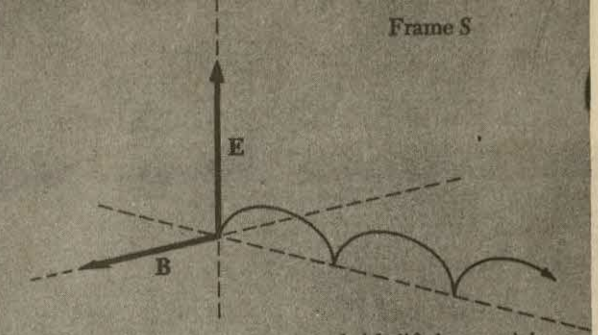
On equating the right-hand sides of (79) and (80) we have

$$\frac{2\pi c^2EM}{qB^2} = \frac{2\pi cMv_1}{qB}, \tag{81}$$

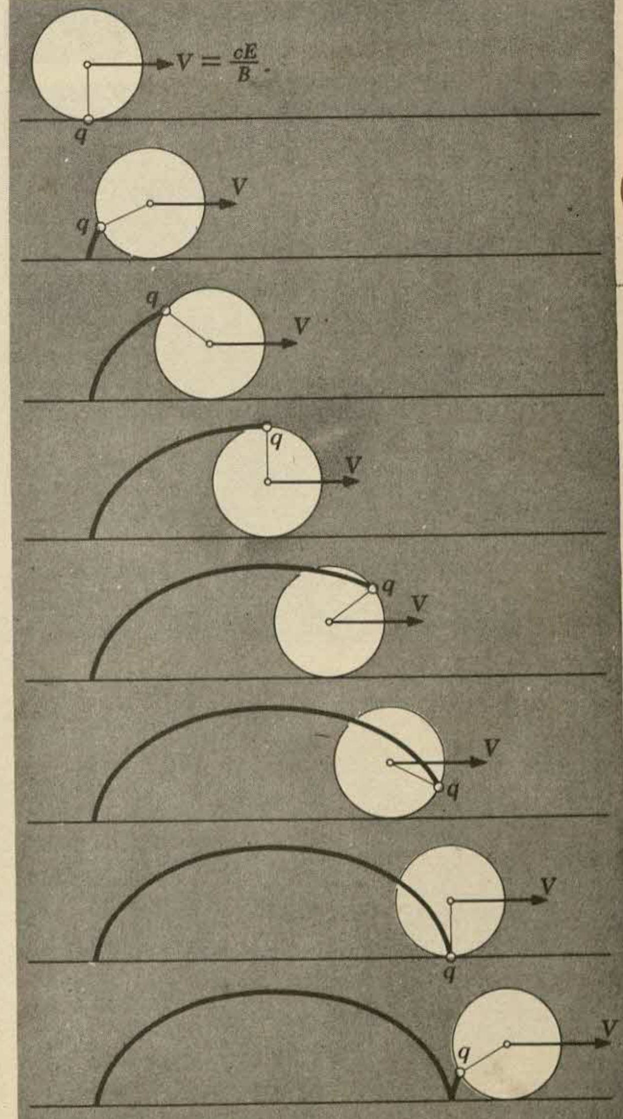
or

$$\frac{cE}{B} = v_1. \tag{82}$$

But $V = cE/B$, so that condition (82) may be rewritten as $V = v_1$. This is exactly what we expect for rolling without slipping, the motion obtained if the particle is started from rest. Only for the special velocity $v_1 = cE/B$ can both $v_x(t)$ and $v_y(t)$ be zero at the same instant. In rolling without slipping, the point on the wheel in contact with the plane is the only point at rest.



The orbit is a common cycloid (if the particle starts from rest), and q has average velocity to the right: $V = \frac{cE}{B}$.



The common cycloid is traced out by q on the circumference of a circle which rolls on a straight line.

Advanced Topic 2. Transformations of the reference frame

When we treated the trajectory of a proton† in crossed electric and magnetic fields, we saw that the motion was particularly simple if viewed from the reference frame moving with the constant velocity

$$\mathbf{V} = \left(0, -\frac{cE}{B}, 0\right) \tag{83}$$

relative to the laboratory reference frame, as in (78). An observer in the new reference frame sees a helical motion. In the frame

$$\mathbf{V} = \left(0, -\frac{cE}{B}, v_z\right), \tag{84}$$

an observer sees a simple circular motion. Here we have added to (83) a velocity along the z axis (which is the direction of $\mathbf{B}$) equal to the constant velocity component of the particle in that direction. By going to the moving system we have simplified enormously the description of the motion of the particle. In the laboratory system the projection of the motion in the xy plane is a cycloid—either a common, prolate, or curtate cycloid,‡ according to whether cE/B is equal to, less than, or greater than the projection $v_\perp$ of the particle velocity on the xy plane. The reference frame in which the motion is circular is clearly the natural frame in which to discuss the motion.

The most powerful single intellectual device known in physics is the transformation of the reference frame from which we observe a process. The essential physics and the essential simplicity of a process are often elicited only after an appropriate transformation of the reference frame. Usually making the transformation is entirely straightforward. The student will have difficulties only if he forgets whether the event§ is to be transformed or the reference frame from which we observe the event. Nearly always at the beginning of this course *the reference frame is to be transformed.* The process or event occurs independently of the reference frame—it is only the *description* of the process relative to a reference frame that is changed when the reference frame is changed. Chapter 3 gave a full discussion of what is assumed in making a transformation of the reference frame. Now we shall undertake some technical practice in making a simple transformation.

Consider a concrete example of the way the description of a process is changed by a change of reference frame. The general motion of a charged particle in crossed electric and magnetic fields is given by (77):

---

† The proton is taken as the particle simply to make a definite choice of the sign of the charge.

‡ A *common cycloid* is the curve traced by a point on the circumference of a circle rolling on a straight line; a *prolate cycloid* is traced by a point on the extension of a radius beyond the circumference; and a *curtate cycloid* is traced by a point lying on a radius but within the circumference.

§ It is fairly easy to recognize when the event (process) is to be transformed; for example, when we say, "Change the sign of all charges, replacing e by $-e$," or, "Reverse the signs of all magnetic fields, replacing $\mathbf{B}$ by $-\mathbf{B}$," or, "Reverse the directions of all velocities, replacing $\mathbf{v}$ by $-\mathbf{v}$."

$$v_x(t) = v_1 \sin \omega_c t; \tag{85a}$$

$$v_y(t) = v_1 \cos \omega_c t - \frac{cE}{B}; \tag{85b}$$

$$v_z(t) = \text{const.} \equiv v_z. \tag{85c}$$

This motion is in the *laboratory system*—that is, in the system in which the magnets and condensers producing the fields **B** and **E** are at rest.

Now change the frame of reference to a frame moving with a uniform velocity with respect to the laboratory and its apparatus. We move to the particular frame having the velocity (84) with respect to the laboratory. In usual vector notation (84) is written

$$\mathbf{V} = -\frac{cE}{B}\hat{\mathbf{y}} + v_z\hat{\mathbf{z}}. \tag{86}$$

We assume that $V/c \ll 1$, so that the problem is nonrelativistic.

We need first of all a clear notation so that we always know which frame we are talking about. Let us agree to denote with a prime a quantity measured in the moving frame. To get from a particle velocity $\mathbf{v}'$ measured in the moving frame to a velocity $\mathbf{v}$ measured in the laboratory frame we add $\mathbf{V}$ to $\mathbf{v}'$:

$$\mathbf{v}' + \mathbf{V} = \mathbf{v}. \tag{87}$$

We must be clear about the sign of **V** here, and this is an easy matter. Suppose the particle is at rest in the moving frame; then $\mathbf{v}' = 0$. But as observed in the laboratory frame the particle has the velocity **V** of the moving frame, in agreement with (87). This type of test is invaluable in straightening out the signs in transformation equations.

With the transformation (87) applied to (85) we have

$$v'_{x'}(t) = v_1 \sin \omega_c t; \tag{88a}$$

$$v'_{y'}(t) = v_1 \cos \omega_c t; \tag{88b}$$

$$v'_{z'}(t) = 0. \tag{88c}$$

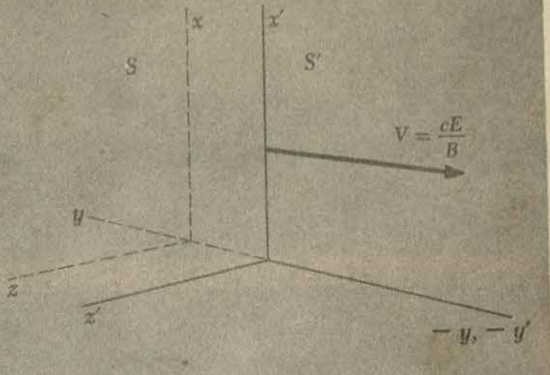
We have been cautious and in $v'_{x'}$ etc., have primed both the v and the x. But no rotation of reference frames is involved, and the x', y', z' axes are parallel respectively to the x, y, z axes. The motion (88) is circular, with angular frequency ω_c, exactly as if in the moving frame the particle were acted on only by a magnetic field. We will return often in Vol. II to the behavior of electric and magnetic fields in moving frames of reference.

For completeness, consider the position coordinates of the particle. In the laboratory frame we integrate (85) to obtain the trajectory:

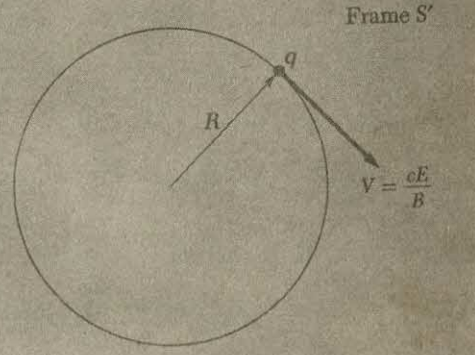
$$x(t) = x_0 - \frac{v_1}{\omega_c}\cos \omega_c t; \tag{89a}$$

$$y(t) = y_0 + \frac{v_1}{\omega_c}\sin \omega_c t - \frac{cEt}{B}; \tag{89b}$$

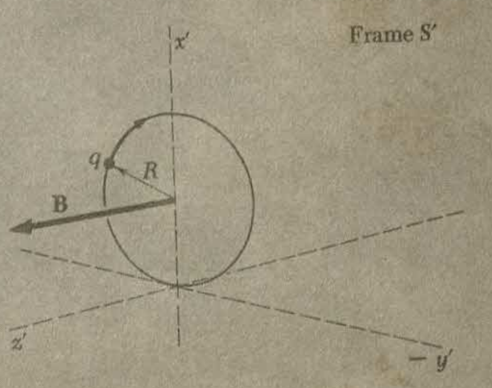
$$z(t) = z_0 + v_z t. \tag{89c}$$



Let us observe q from reference frame S', obtained from frame S by a Galilean transformation.



In S', q describes a circle with speed V and radius $R = \frac{MVc}{EB}$.



Thus in S', $\mathbf{E} = 0$ and there is only a magnetic field **B**.

In the moving frame we integrate (88) to obtain the trajectory:

$$x'(t) = x_0' - \frac{v_1}{\omega_c}\cos\omega_c t; \tag{90a}$$

$$y'(t) = y_0' + \frac{v_1}{\omega_c}\sin\omega_c t; \tag{90b}$$

$$z'(t) = z_0'. \tag{90c}$$

The trajectory in the primed frame is a circle of radius $\rho = v_1/\omega_c$. In order to determine x_0', y_0', z_0' in terms of x_0, y_0, z_0 we need to know the distance between the origins 0, 0′ of the two reference frames at $t = 0$. If the origins are coincident at $t = 0$, then $(x_0', y_0', z_0') = (x_0, y_0, z_0)$.

We have now viewed the same process in two frames of reference. In the laboratory frame the proton moves in a complicated cycloidal orbit under the influence of crossed **E** and **B** fields:

$$M\frac{d^2\mathbf{r}}{dt^2} = e\mathbf{E} + \frac{e}{c}\mathbf{v}\times\mathbf{B}. \tag{91}$$

In the single-prime frame, which moves with the particular uniform velocity **V** with respect to the laboratory frame, the particle moves as if only a magnetic field were present:

$$M\frac{d^2\mathbf{r}'}{dt^2} = \frac{e}{c}\mathbf{v}'\times\mathbf{B}', \tag{92}$$

where we have assumed that $\mathbf{B}' = \mathbf{B}$ in the nonrelativistic region.

Mathematical Note. Complex numbers

The mathematical discussion of oscillatory motion and in particular the analysis of ac circuits is facilitated by the use of complex numbers. A complex number z may be written

$$z = a + ib, \tag{93}$$

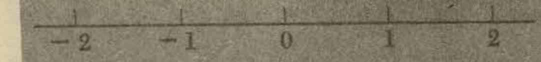
where a and b are ordinary real numbers, and i is the square root of minus one:

$$i = \sqrt{-1}; \qquad i^2 = -1. \tag{94}$$

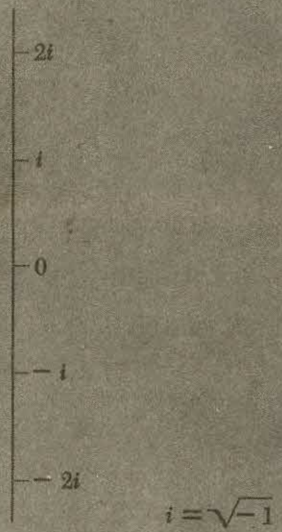
The real part of z is written $\mathrm{Re}(z)$ and is equal to a. The imaginary part is written $\mathrm{Im}(z)$ and is equal to b. This is a matter of convention: it is superfluous to say $\mathrm{Im}(z) = ib$, so we say $\mathrm{Im}(z) = b$. Thus $z = \mathrm{Re}(z) + i\,\mathrm{Im}(z)$.

To *add* two complex numbers, form

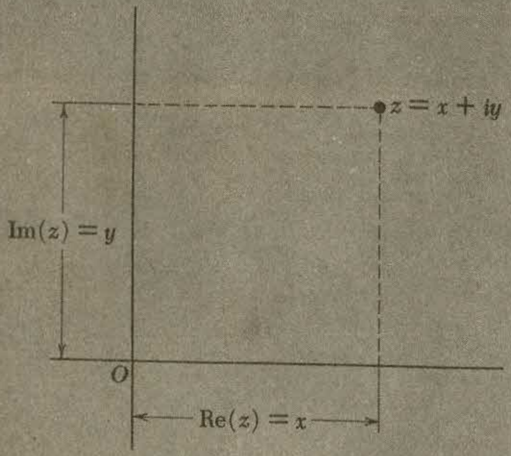
$$z_1 + z_2 = (a_1 + ib_1) + (a_2 + ib_2) = (a_1 + a_2) + i(b_1 + b_2). \tag{95}$$



The real numbers are readily represented by points on the horizontal axis, . . .



and the imaginary numbers by points on the vertical axis.



A *complex* number $z = x + iy$ is represented by a point in the "complex" or z plane. Note that the number $+i$ is represented by a distance $+1$ in the y direction.

To *multiply* two complex numbers, form

$$z_1z_2 = (a_1 + ib_1)(a_2 + ib_2) = a_1a_2 + ia_1b_2 + ia_2b_1 + i^2b_1b_2; \qquad (96)$$

using $i^2 = -1$, we have

$$z_1z_2 = (a_1a_2 - b_1b_2) + i(a_1b_2 + a_2b_1). \qquad (97)$$

To *divide,* we usually want to manipulate the quotient so that the denominator will be real:

$$\frac{z_1}{z_2} = \frac{a_1 + ib_1}{a_2 + ib_2} = \frac{(a_1 + ib_1)(a_2 - ib_2)}{(a_2 + ib_2)(a_2 - ib_2)}$$
$$= \frac{(a_1a_2 + b_1b_2) + i(a_2b_1 + a_1b_2)}{a_2^2 + b_2^2}. \qquad (98)$$

The *complex conjugate* z^* of any complex number is defined by changing the sign of i wherever it appears. If $z = a + ib$, then

$$z^* = a - ib. \qquad (99)$$

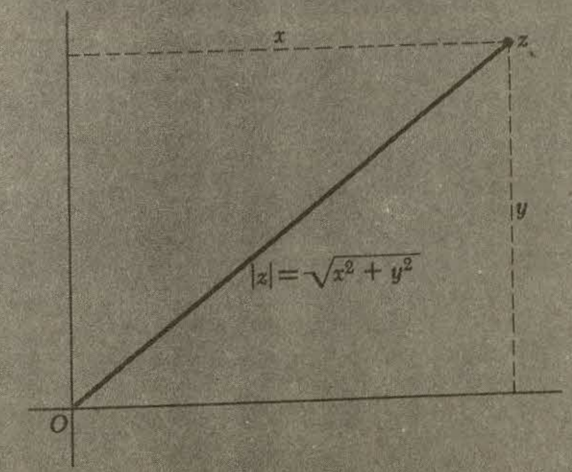
The product of a complex number and its conjugate is a positive real number:

$$zz^* = a^2 + b^2. \qquad (100)$$

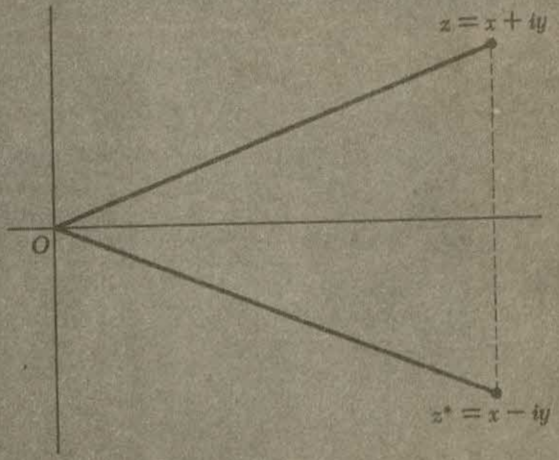
We define the *modulus* of z to be

$$|z| = \sqrt{zz^*} = \sqrt{a^2 + b^2}.$$

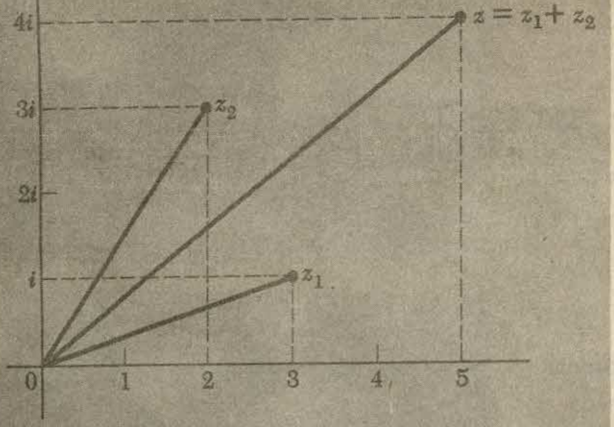
These definitions become more intuitive if we represent the complex number $z = x + iy$ geometrically. Let the x axis be the real axis, and the



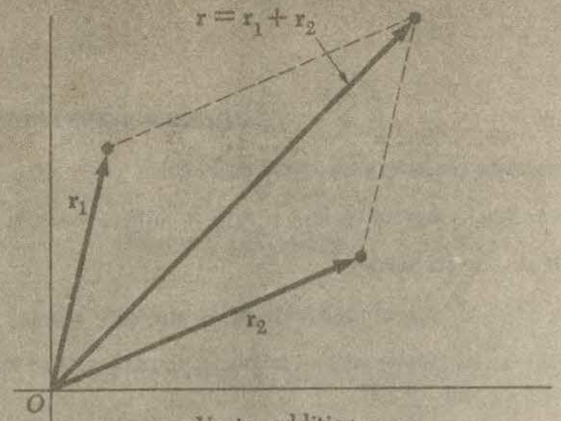
The absolute value $|z|$ of the number z is the distance from the origin to the point z. $|z| = \sqrt{x^2 + y^2}$ by the Pythagorean theorem.



The complex conjugate of $z = x + iy$
is $z^* = x - iy$.
Obviously $|z| = |z^*|$.

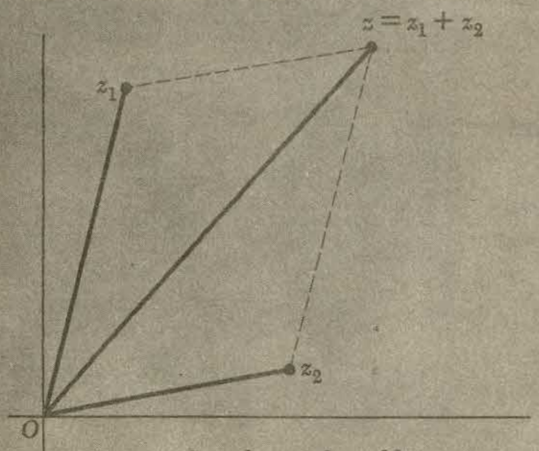


Addition of complex numbers:
If $z_1 = (x_1 + iy_1)$ and $z_2 = (x_2 + iy_2)$, then $z = z_1 + z_2 = (x_1 + x_2) + i(y_1 + y_2)$.
See that this is verified in the illustration.



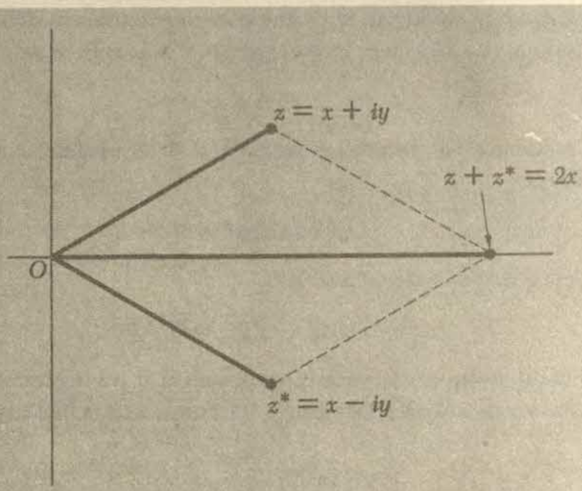
Vector addition

Vectors also add "component by component." Thus, because the parallelogram rule holds for vector addition, . . .

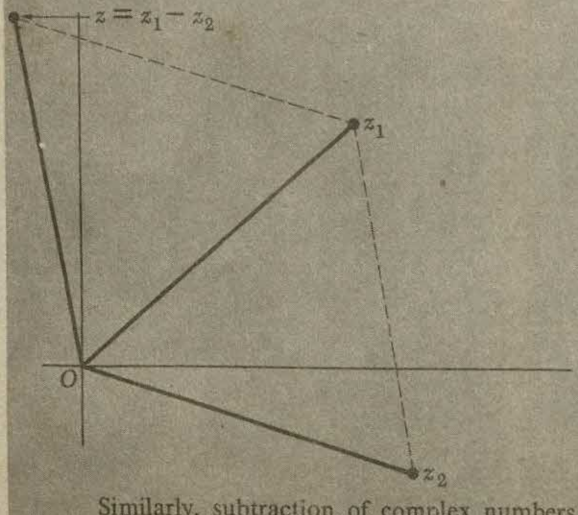


Complex number addition

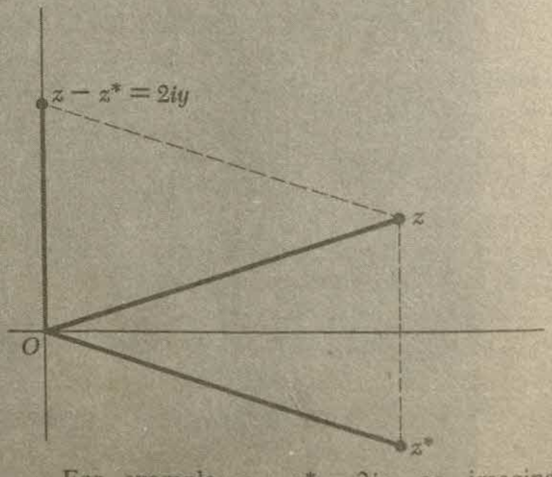
it holds for addition of complex numbers as well.



For example, $z + z^* = 2x$, a real number.



Similarly, subtraction of complex numbers is easily obtained by the parallelogram rule . . .



For example, $z - z^* = 2iy$, an imaginary number.

y axis be the imaginary axis. Then $|z|$ is the distance from the origin 0 to z, and $\mathrm{Re}(z)$, $\mathrm{Im}(z)$ are the projections of the line $0z$ on the two axes. In the graphical representation, the addition of complex numbers obeys the parallelogram law; this result is confirmed analytically. Complex numbers have some of the properties of vectors in a two-dimensional space.

Let φ be the angle between $0z$ and the real axis. Then we may represent any complex number $z = x + iy$ in the form

$$z = |z|(\cos\varphi + i\sin\varphi), \tag{101}$$

where

$$x = |z|\cos\varphi \qquad \text{and} \qquad y = |z|\sin\varphi. \tag{102}$$

We introduce the important relation connecting the exponential, the cosine, and the sine:

$$e^{i\varphi} = \cos\varphi + i\sin\varphi, \tag{103}$$

which follows from the series expansion of the three terms; then (101) may be written as

$$z = |z|e^{i\varphi}. \tag{104}$$

We say that φ is the *argument* of z, and $|z|$ is the *modulus* of z. This form is particularly suited to the representation of the amplitude and phase of an oscillation. Multiplication and division of complex numbers now appear in the simpler forms

$$z_1z_2 = |z_1||z_2|e^{i(\varphi_1+\varphi_2)}; \tag{105}$$

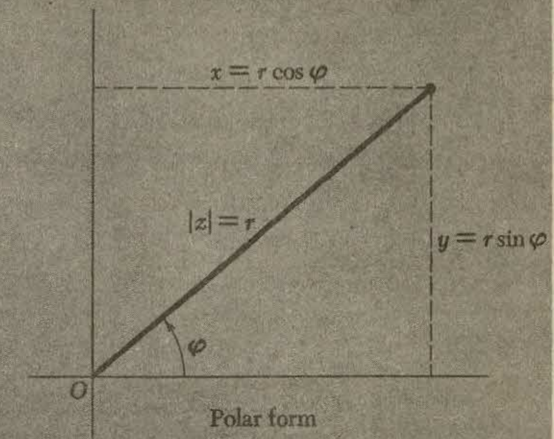
$$\frac{z_1}{z_2} = \frac{|z_1|}{|z_2|}e^{i(\varphi_1-\varphi_2)}. \tag{106}$$

These forms are well-suited to numerical calculation. From the graphical representation of a complex number as a vector, we see that multiplying it by another complex number is equivalent to stretching the vector and rotating it. The results are independent of the order of the operations.

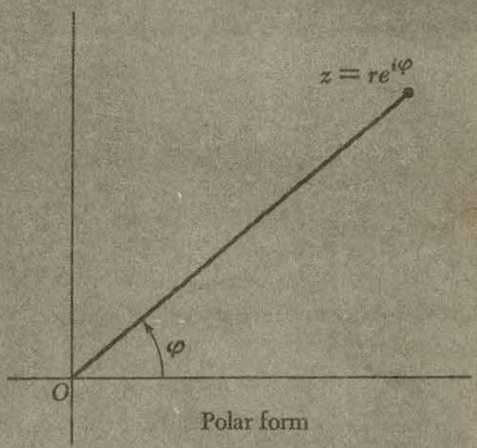
The complex numbers of modulus unity ($|z| = 1$) lie on the unit circle about the origin of the complex plane defined by $z = x + iy$ and are of the form (103). Using (103) and (105), we derive the de Moivre equation:

$$(\cos\varphi + i\sin\varphi)^n = (e^{i\varphi})^n = e^{in\varphi} = \cos n\varphi + i\sin n\varphi, \tag{107}$$

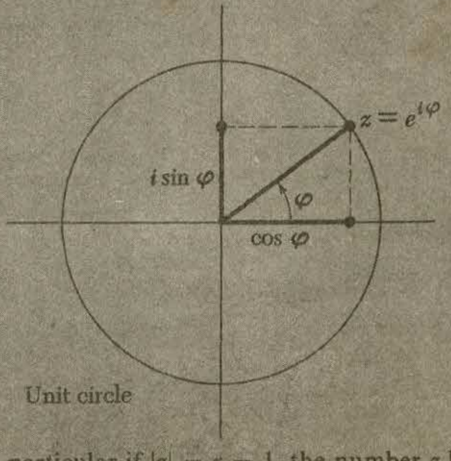
which, by separately equating real and imaginary parts, gives us directly the equations for the sine and cosine of multiple angles.



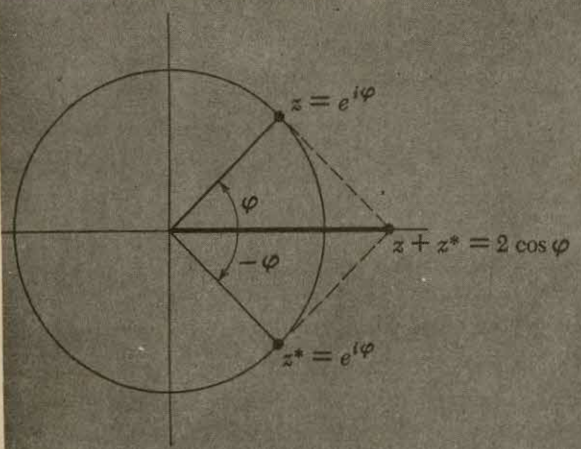
Clearly $z = x + iy = r(\cos\varphi + i\sin\varphi)$. Since $\cos\varphi + i\sin\varphi = e^{i\varphi}$, . . .



we can write any complex number in *polar form:* $z = re^{i\varphi}$.

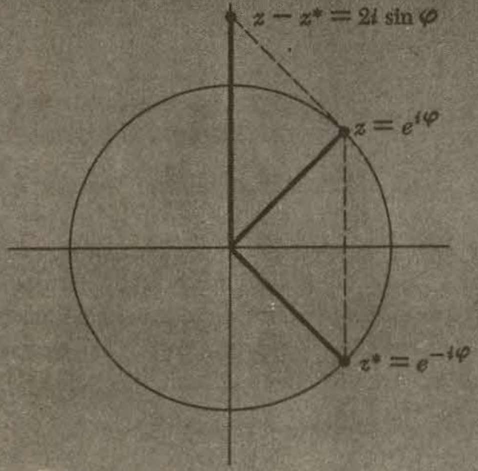


In particular if $|z| = r = 1$, the number z lies on the unit circle. The polar form is very convenient . . .



As we saw before, $z + z^* = 2x$. Thus $e^{i\varphi} + e^{-i\varphi} = 2\cos\varphi$,.

or $$\boxed{\cos\varphi = \frac{e^{i\varphi} + e^{-i\varphi}}{2}.}$$



Similarly, $z - z^* = 2iy$, so

$$\boxed{\sin\varphi = \frac{1}{2i}(e^{i\varphi} - e^{-i\varphi}).}$$

Example. Find the equations for $\sin 2\varphi$ and $\cos 2\varphi$. From the de Moivre equation with $n = 2$,

$$(e^{i\varphi})^2 = (\cos\varphi + i\sin\varphi)^2 = e^{2i\varphi} = \cos 2\varphi + i\sin 2\varphi, \tag{108}$$

or

$$\cos^2\varphi - \sin^2\varphi + 2i\cos\varphi\sin\varphi = \cos 2\varphi + i\sin 2\varphi. \tag{109}$$

On equating real and imaginary parts on both sides of (109), we have the results

$$\cos^2\varphi - \sin^2\varphi = \cos 2\varphi; \tag{110}$$

$$2\cos\varphi\sin\varphi = \sin 2\varphi. \tag{111}$$

A great deal of trigonometry follows naturally and easily on the use of the de Moivre equation.

As an exercise use (103) to establish the important identities

$$\cos\varphi = \frac{e^{i\varphi} + e^{-i\varphi}}{2}, \qquad \sin\varphi = \frac{e^{i\varphi} - e^{-i\varphi}}{2i}. \tag{112}$$

An equation in complex numbers must be satisfied separately by its real and imaginary parts. Thus we may work with a real oscillation $\Psi = \cos\omega t$ in the complex form $\Psi = e^{-i\omega t}$, and extract the real part when we are finished. We may do this freely when products of complex numbers are not involved—that is, provided the equations are linear in complex numbers. But *great caution is needed with products:* Suppose that we are interested in the product x_1x_2 of two real quantities. If we write

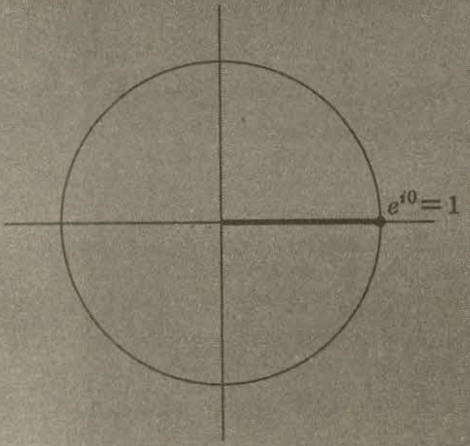
$$z_1 = x_1 + iy_1; \qquad z_2 = x_2 + iy_2, \tag{113}$$

the real part of the product is

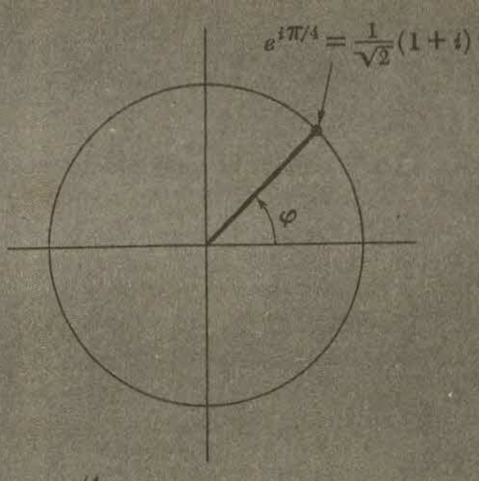
$$\text{Re}(z_1z_2) = x_1x_2 - y_1y_2, \tag{114}$$

and this is *not equal* to the product of the real parts:

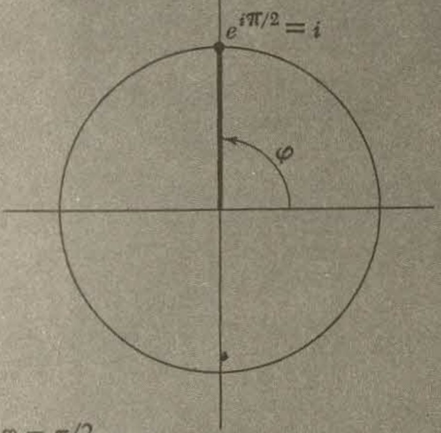
$$\text{Re}(z_1)\,\text{Re}(z_2) = x_1x_2. \tag{115}$$



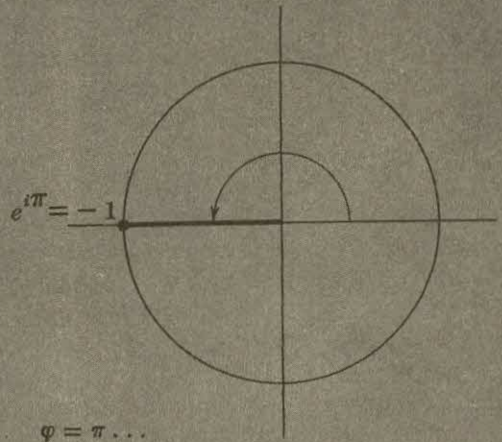
Let us look at complex numbers on the unit circle. Here $\varphi = 0$, . . .



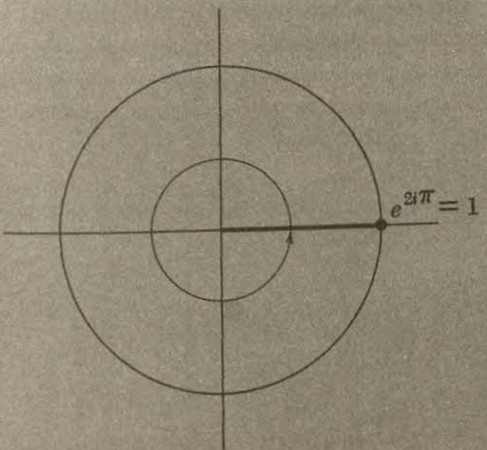
$\varphi = \pi/4$, . . .



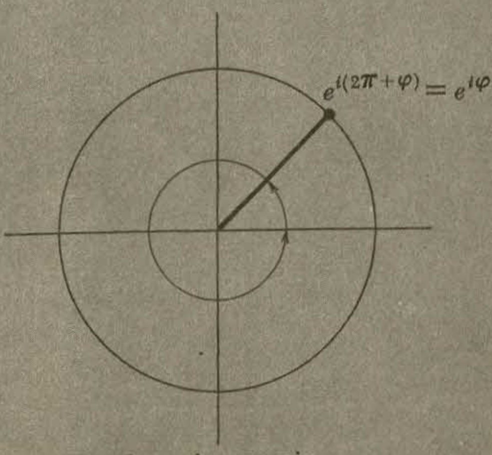
$\varphi = \pi/2$, . . .



$\varphi = \pi$ . . .



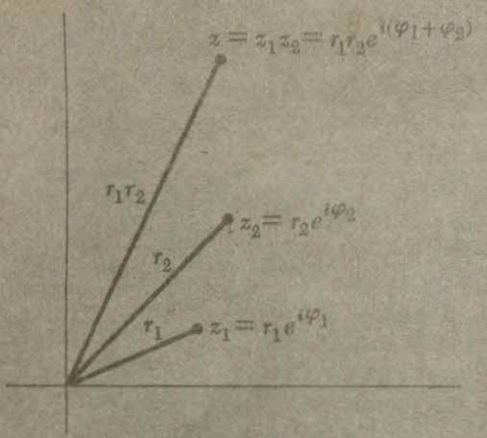
Now $\varphi = 2\pi$. We have returned to the starting point.



For $\varphi > 2\pi$ we have, again,

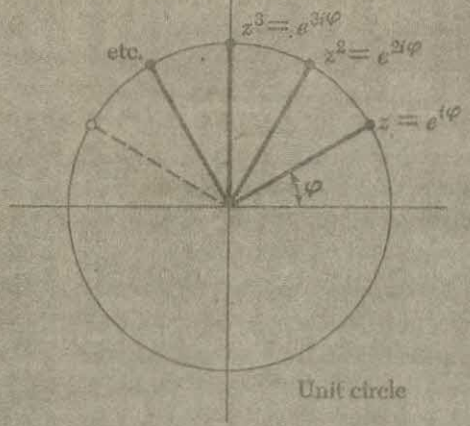
$$e^{i(2\pi+\varphi)} = e^{i\varphi}.$$

Thus $e^{i\varphi}$ is a *periodic function with period* 2π.

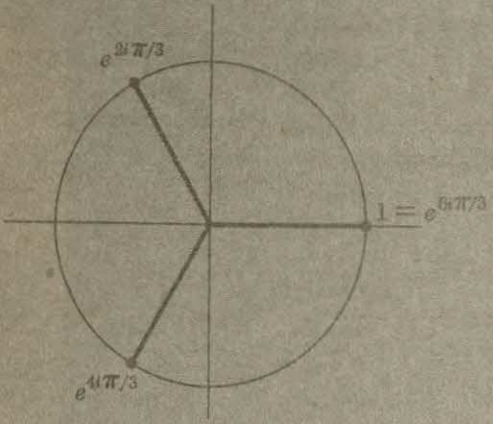

$z = z_1 z_2 = r_1 r_2 e^{i(\varphi_1+\varphi_2)}$
$r_1 r_2$
$z_2 = r_2 e^{i\varphi_2}$
r_2
r_1
$z_1 = r_1 e^{i\varphi_1}$

If $z_1 = r_1 e^{i\varphi_1}$ and $z_2 = r_2 e^{i\varphi_2}$, then

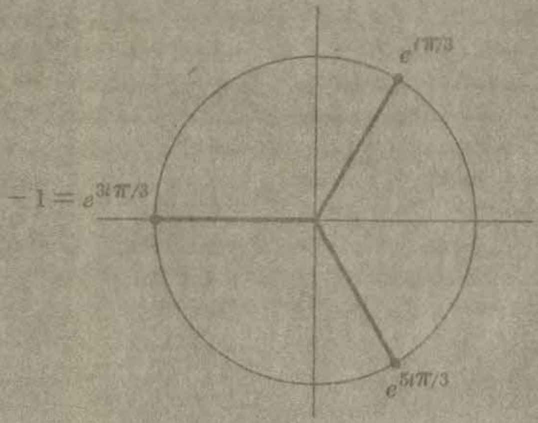
$$z = z_1 z_2 = r_1 r_2 e^{i(\varphi_1+\varphi_2)}.$$



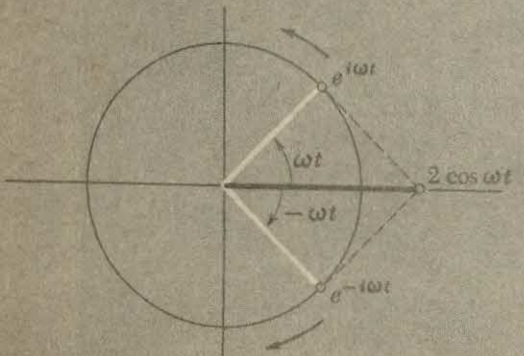
For example, the successive powers of $z = e^{i\varphi}$ are as shown.



Another example: Here are the three roots of the polynomial $f(z) = z^3 - 1, \ldots$



and here are the three roots of the polynomial $g(z) = z^3 + 1$. Verify the last two examples.



Complex numbers play an important role in the theory of oscillations: An oscillation $\cos \omega t$ can be expressed as a linear combination of two oppositely rotating complex numbers:

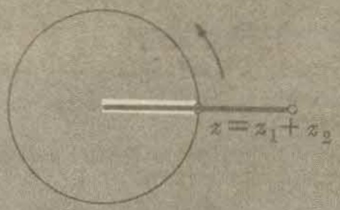
$$\cos \omega t = \left(\frac{e^{i\omega t} + e^{-i\omega t}}{2}\right).$$

Although complex notation is very convenient in solving linear differential equations and in analyzing linear circuits, it must be handled with care when calculating bilinear quantities such as the power absorption and the energy flux. For this reason the laboratory manual for this course makes relatively little use of complex numbers. But without complex numbers, quantum physics would be quite clumsy.

Example. Find the resultant of the oscillations $\psi_1 = \cos \omega t$ and $\psi_2 = \cos(\omega + \Delta\omega)t$. In complex form

$$\begin{aligned}\psi_1 + \psi_2 &= e^{-i\omega t} + e^{-i(\omega+\Delta\omega)t} \\ &= \left(e^{i(\Delta\omega t/2)} + e^{-i(\Delta\omega t/2)}\right)e^{-i(\omega+\Delta\omega/2)t} \\ &= 2(\cos \Delta\omega t/2)e^{-i(\omega+\Delta\omega/2)t}, \qquad (116)\end{aligned}$$

using the identity (112). Thus



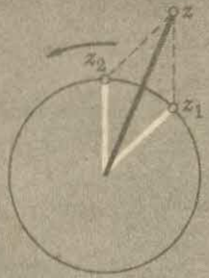
As another example, two complex numbers

$$z_1 = e^{i(\omega-\Delta\omega/2)t}$$

and

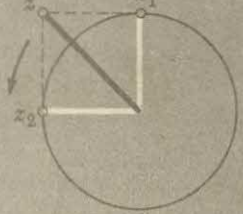
$$z_2 = e^{i(\omega+\Delta\omega/2)t}$$

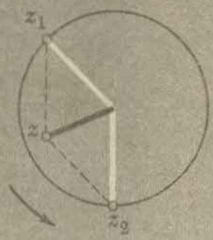
are in phase at $t = 0$. However z_2 rotates faster than z_1.



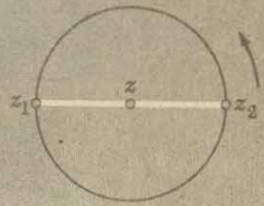
As time progresses, z_1 and z_2 get further out of phase. Their resultant: $z = z_1 + z_2$ rotates with frequency ω:

$$z = z_0 e^{i\omega t}.$$



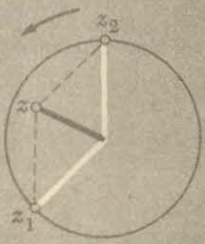


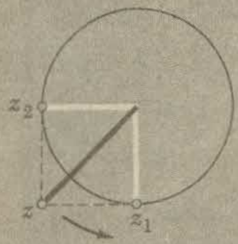
The real amplitude z_0 varies with time.

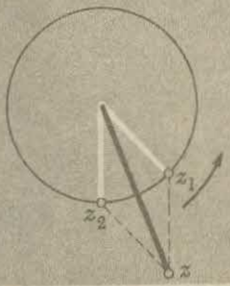


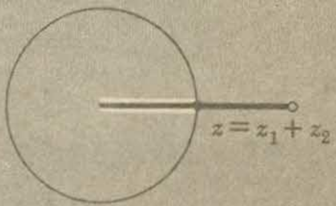
When $\Delta\omega t = \pi$,

$$z_0 = 0.$$









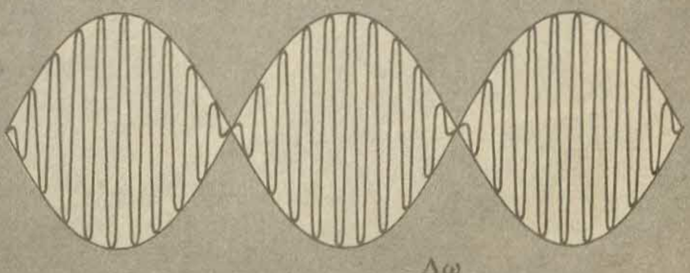
Thus, since

$$\begin{aligned} z = z_1 + z_2 &= e^{i(\omega+\Delta\omega/2)t} + e^{i(\omega-\Delta\omega/2)t} \\ &= 2\left(\cos\frac{\Delta\omega t}{2}\right)e^{i\omega t}, \end{aligned}$$

we have

$$z_0 = 2\cos\frac{\Delta\omega t}{2}.$$

Thus $\mathrm{Re}(z) = 2\cos\frac{\Delta\omega}{2}t$ looks like this if $\omega \gg \Delta\omega$.



$$\cos \omega t + \cos(\omega + \Delta\omega)t = \mathrm{Re}(\psi_1 + \psi_2)$$
$$= 2 \cos \frac{\Delta\omega t}{2} \cos\left(\omega + \frac{\Delta\omega}{2}\right)t. \quad (117)$$

If $\Delta\omega/\omega$ is small compared with unity ($\Delta\omega/\omega \ll 1$), the superposition of the two oscillations produces a slow modulation of the fundamental frequency ω.

Example. Consider the coupled differential equations:

$$\dot{v}_x = \omega_c v_y; \qquad \dot{v}_y = -\omega_c v_x. \quad (118)$$

These describe the motion of a charged particle in a uniform magnetic field directed along the z axis. Multiply the second of these equations by $-i$ and add it to the first to obtain

$$\dot{v}_x - i\dot{v}_y = \omega_c(v_y + iv_x). \quad (119)$$

We define

$$v^+ \equiv v_x + iv_y; \qquad v^- \equiv v_x - iv_y, \quad (120)$$

whence (119) becomes

$$\frac{dv^-}{dt} = i\omega_c v^-, \quad (121)$$

which has the solution

$$v^- = Ae^{i(\omega_c t + \varphi)}, \quad (122)$$

where A and φ are constants. On taking real and imaginary parts,

$$\begin{aligned} v_x &= \mathrm{Re}(v^-) = A \cos(\omega_c t + \varphi); \\ v_y &= -\mathrm{Im}(v^-) = -A \sin(\omega_c t + \varphi). \end{aligned} \quad (123)$$

Problems on complex numbers

1. (*a*) If $z_1 = 5 + 3i$ and $z_2 = 5i$, what is $z_1 + z_2$?
    (*b*) What is $z_1 z_2$?
    (*c*) What is $z_1 + z_1^*$?
    (*d*) What is $z_1 - z_1^*$?
    (*e*) What is $z_1 z_1^*$?
2. Express all the answers to Prob. 1 in the polar form $|z|e^{i\varphi}$.
3. Show by squaring both sides that $\sqrt{i} = \dfrac{1 + i}{\sqrt{2}}$.
4. What is the square root of $z = 4 + 9i$?

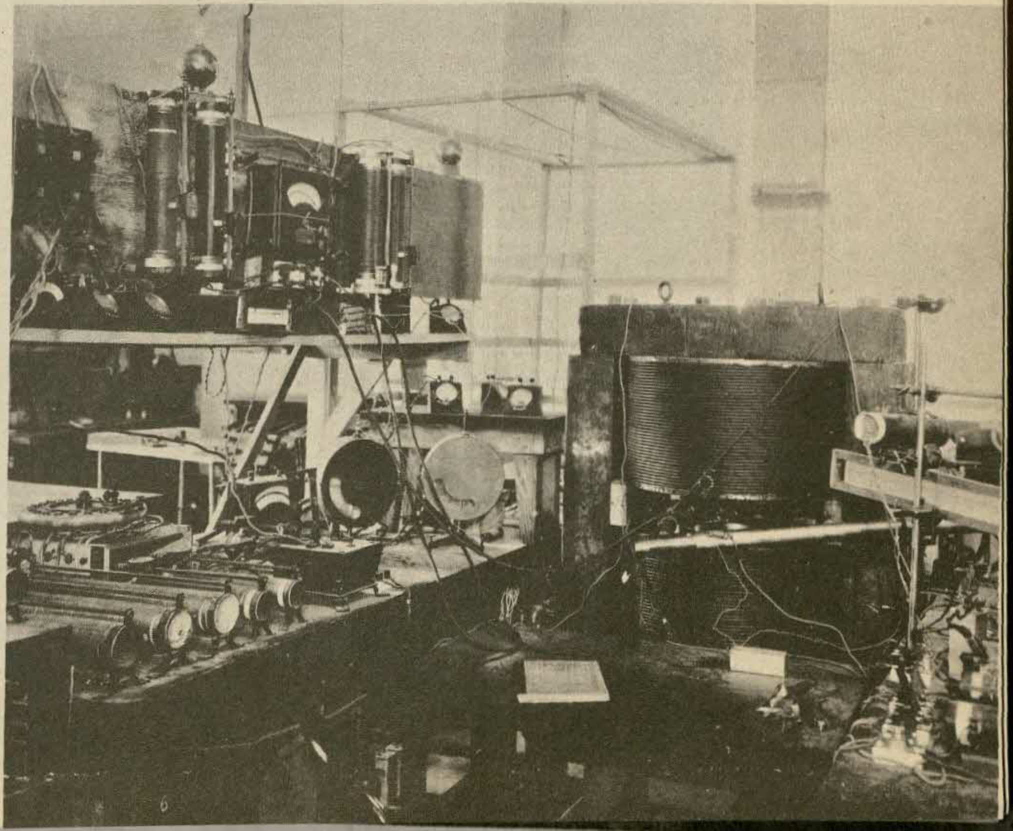
Historical Note. Invention of the cyclotron

Most of the present high-energy particle accelerators are descended from the first 1-Mev proton cyclotron built by E. O. Lawrence and M. S. Livingston in LeConte Hall at Berkeley. The cyclotron was conceived by Lawrence; the conception was first published by Lawrence and Edlefsen in a talk abstracted in Science **72**, 376, 377 (1930). In 1932 the first results were published in a beautiful paper in the *Physical Review*, principal physics journal of the American Physical Society. Although this journal requires all papers to be accompanied by informative abstracts, few are so clear and informative as that reproduced here from the classic paper by Lawrence and Livingston. Also reproduced are two figures from the original paper. Professor Livingston is at M.I.T.; Professor Lawrence died in 1958.

The original 11-in. magnet was almost immediately outgrown for accelerator applications; it has been rebuilt and is still used for a variety of research projects in LeConte Hall. The first successful experiments on cyclotron resonance of charge carriers in crystals were carried out with this magnet.

For an interesting account of the early history of the cyclotron, see E. O. Lawrence, "The Evolution of the Cyclotron," *Les Prix Nobel en 1951*, pp. 127–140 (Imprimerie Royale, Stockholm, 1952).

An early cyclotron.



APRIL 1, 1932 PHYSICAL REVIEW VOLUME 40

THE PRODUCTION OF HIGH SPEED LIGHT IONS WITHOUT THE USE OF HIGH VOLTAGES

By Ernest O. Lawrence and M. Stanley Livingston
University of California

(Received February 20, 1932)

Abstract

The study of the nucleus would be greatly facilitated by the development of sources of high speed ions, particularly protons and helium ions, having kinetic energies in excess of 1,000,000 volt-electrons; for it appears that such swiftly moving particles are best suited to the task of nuclear excitation. The straightforward method of accelerating ions through the requisite differences of potential presents great experimental difficulties associated with the high electric fields necessarily involved. The present paper reports the development of a method that avoids these difficulties by means of the multiple acceleration of ions to high speeds without the use of high voltages. The method is as follows: Semi-circular hollow plates, not unlike duants of an electrometer, are mounted with their diametral edges adjacent, in a vacuum and in a uniform magnetic field that is normal to the plane of the plates. High frequency oscillations are applied to the plate electrodes producing an oscillating electric field over the diametral region between them. As a result during one half cycle the electric field accelerates ions, formed in the diametral region, into the interior of one of the electrodes, where they are bent around on circular paths by the magnetic field and eventually emerge again into the region between the electrodes. The magnetic field is adjusted so that the time required for traversal of a semi-circular path within the electrodes equals a half period of the oscillations. In consequence, when the ions return to the region between the electrodes, the electric field will have reversed direction, and the ions thus receive second increments of velocity on passing into the other electrode. Because the path radii within the electrodes are proportional to the velocities of the ions, the time required for a traversal of a semi-circular path is independent of their velocities. Hence if the ions take exactly one half cycle on their first semi-circles, they do likewise on all succeeding ones and therefore spiral around in resonance with the oscillating field until they reach the periphery of the apparatus. Their final kinetic energies are as many times greater than that corresponding to the voltage applied to the electrodes as the number of times they have crossed from one electrode to the other. This method is primarily designed for the acceleration of light ions and in the present experiments particular attention has been given to the production of high speed protons because of their presumably unique utility for experimental investigations of the atomic nucleus. Using a magnet with pole faces 11 inches in diameter, a current of 10^{-9} ampere of 1,220,000 volt-protons has been produced in a tube to which the maximum applied voltage was only 4000 volts. There are two features of the developed experimental method which have contributed largely to its success. First there is the focussing action of the electric and magnetic fields which prevents serious loss of ions as they are accelerated. In consequence of this, the magnitudes of the high speed ion currents obtainable in this indirect manner are comparable with those conceivably obtainable by direct high voltage methods. Moreover, the focussing action results in the generation of very narrow beams of ions—less than 1 mm cross-sectional diameter—which are ideal for experimental studies of collision processes. Of hardly less importance is the second feature of the method which is the simple and highly effective means for the correction of the magnetic field along the paths of the ions. This makes it possible, indeed easy, to operate the tube effectively with a very high amplification factor (i.e., ratio of final equivalent voltage of accelerated ions to applied voltage). In consequence, this method in its present stage of development constitutes a highly reliable and experimentally convenient source of high speed ions requiring relatively modest laboratory equipment. Moreover, the present experiments indicate that this indirect method of multiple acceleration now makes practicable the production in the laboratory of protons having kinetic energies in excess of 10,000,000 volt-electrons. With this in mind, a magnet having pole faces 114 cm in diameter is being installed in our laboratory.

INTRODUCTION

THE classical experiments of Rutherford and his associates[1] and Pose[2] on artificial disintegration, and of Bothe and Becker[3] on excitation of nuclear radiation, substantiate the view that the nucleus is susceptible to the same general methods of investigation that have been so successful in revealing the extra-nuclear properties of the atom. Especially do the results of their work point to the great fruitfulness of studies of nuclear transitions excited artificially in the laboratory. The development of methods of nuclear excitation on an extensive scale is thus a problem of great interest; its solution is probably the key to a new world of phenomena, the world of the nucleus.

But it is as difficult as it is interesting, for the nucleus resists such experimental attacks with a formidable wall of high binding energies. Nuclear energy levels are widely separated and, in consequence, processes of nuclear excitation involve enormous amounts of energy—millions of volt-electrons.

It is therefore of interest to inquire as to the most promising modes of nuclear excitation. Two general methods present themselves; excitation by absorption of radiation (gamma radiation), and excitation by intimate nuclear collisions of high speed particles.

Of the first it may be said that recent experimental studies [4,5] of the absorption of gamma radiation in matter show, for the heavier elements, variations with atomic number that indicate a quite appreciable nuclear effect. This suggests that nuclear excitation by absorption of radiation is perhaps a not infrequent process, and therefore that the development of an intense artificial source of gamma radiation of various wave-lengths would be of considerable value for nuclear studies. In our laboratory, as elsewhere, this being attempted.

But the collision method appears to be even more promising, in consequence of the researches of Rutherford and others cited above. Their pioneer investigations must always be regarded as really great experimental achievements, for they established definite and important information about nuclear processes of great rarity excited by exceedingly weak beams of bombarding particles—alpha-particles from radioactive sources. Moreover, and this is the point to be emphasized here, their work has shown strikingly the

---

[1] See Chapter 10 of Radiations from Radioactive Substances by Rutherford, Chadwick and Ellis.

[2] H. Pose, Zeits. f. Physik **64**, 1 (1930).

[3] W. Bothe and H. Becker, Zeits. f. Physik **66**, 1289 (1930).

[4] G. Beck, Naturwiss. **18**, 896 (1930).

[5] C. Y. Chao, Phys. Rev. **36**, 1519 (1930).

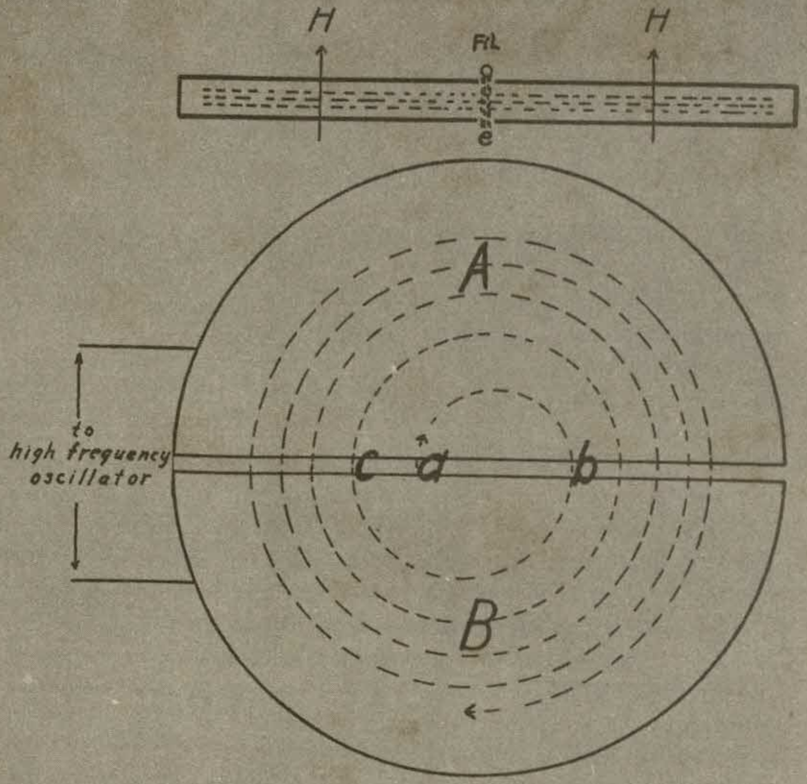


Fig. 1. Diagram of experimental method for multiple acceleration of ions.

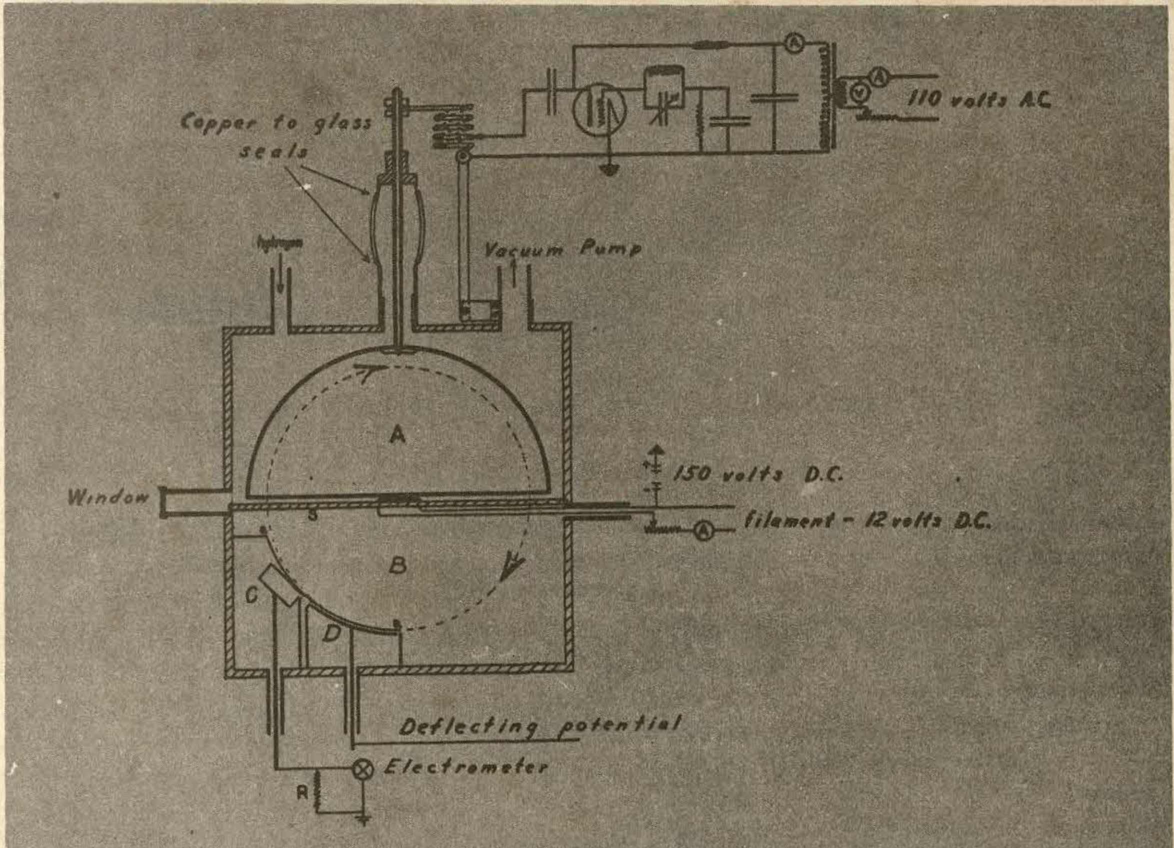


Fig. 2. Diagram of apparatus for the multiple acceleration of ions.

Chapter 5

Conservation laws in the physical world 133
Definition of concepts 134

Conservation of energy 140
Work 141
Kinetic energy 141
Examples: Free fall 142
Linear restoring force 143
Power 144
Conservative forces 145
Potential energy 147
Potential energy in an electric field 151
Examples: Electrostatic field and potential; potential difference; statvolts; volts 153
Electron volts 154
Linear restoring force 155
Escape velocity from earth and from solar system 156
Gravitational potential near surface of the earth 157

Further reading 159

Film list 159

Problems 159

Historical Notes: Discovery of Ceres and of Neptune 162

Conservation of Energy

Conservation Laws in the Physical World. In the physical world there exist a number of conservation laws, some exact and some approximate. A conservation law is usually the consequence of some underlying symmetry in the universe. There are conservation laws relating to energy, linear momentum, angular momentum, charge, number of baryons (protons, neutrons, and heavier elementary particles), strangeness, and various other quantities. We discuss the conservation of energy in this chapter. In the following chapter we discuss the conservation of linear and angular momentum. The entire discussion at present will be phrased for the nonrelativistic regime, which means a restriction to Galilean transformations, speeds very much less than that of light, and independence of mass and energy. In Chap. 12 after we have discussed the Lorentz transformation and special relativity, we shall give the appropriate forms of the energy and momentum conservation laws in the relativistic regime.

If all the forces in a problem are known, and if we are clever enough and have computers of adequate speed and capacity to solve for the trajectories of all the particles, then the conservation laws give us no additional information. But they are very powerful tools which a physicist uses every day. Why are conservation laws powerful tools?

1. Conservation laws are independent of the details of the trajectory and, often, of the details of the particular force. The laws are therefore a way of stating very general and significant consequences of the equations of motion. A conservation law can

sometimes assure us that something is impossible. Thus we do not waste time analyzing an alleged perpetual motion device if it is merely a closed system of mechanical and electrical components, or a satellite propulsion scheme which purportedly works by moving internal weights.

2. Conservation laws have been used even when the force is unknown; this applies particularly in the physics of elementary particles.

3. Conservation laws have an intimate connection with invariance. In the exploration of new and not yet understood phenomena the conservation laws are often the most striking physical fact we know. They may suggest appropriate invariance concepts. In Chap. 3 we saw that the conservation of linear momentum could be interpreted as a direct consequence of the principle of Galilean invariance.

4. Even when the force is known exactly, a conservation law may be a convenient aid in solving for the motion of a particle. Many physicists have a regular routine for solving unknown problems: First we use the relevant conservation laws one by one; only after this, if there is anything left to the problem, will we get down to real work with differential equations, variational and perturbation methods, computers, intuition, and the other tools at our disposal. In Chaps. 7 and 9 we exploit the energy and momentum conservation laws in this way.

Definition of Concepts. The law of conservation of energy involves the concepts of kinetic energy, potential energy, and work. These concepts, which can be understood from a simple example, will be treated in more detail later. To start with, we discuss forces and motions in only one dimension. This simplifies the notation. Most matters in this chapter are treated twice; the repetition may be helpful.

Consider a particle of mass M floating in intergalactic space and free of all external interactions. We observe the particle from an inertial reference frame. A force $\mathbf{F}_{\text{app}}$ is applied to the particle at time $t = 0$. The force thereafter is kept constant in magnitude and direction. The direction is taken to be the x direction. The particle will accelerate under the action of the applied force. The motion at times $t > 0$ is described by Newton's second law:

$$F_{\text{app}} = M\ddot{x}. \qquad (1)$$

Thus the velocity after time t is

$$v(t) = \int_0^t \ddot{x}\, dt = v_0 + \frac{F_{\text{app}}}{M} t, \tag{2}$$

where v_0 is the initial velocity, supposed to lie in the x direction. Notice that (2) may be written as

$$F_{\text{app}} t = Mv(t) - Mv_0. \tag{2a}$$

The right-hand side is the change of momentum of the particle between 0 and t; the left-hand side is called the *impulse* of force between the same times. Equation (2a) tells us that the change in the momentum is equal to the impulse. If the initial position is x_0, the position at time t is given by

$$x(t) = \int_0^t v(t)\, dt = x_0 + v_0 t + \frac{1}{2} \frac{F_{\text{app}}}{M} t^2, \tag{3}$$

on integrating (2) with respect to the time.

We may solve (2) for the time t:

$$t = \frac{M}{F_{\text{app}}}(v - v_0). \tag{4}$$

Now substitute (4) in (3) to obtain

$$\begin{aligned} x - x_0 &= \frac{M}{F_{\text{app}}}(vv_0 - v_0^2) + \frac{1}{2}\frac{M}{F_{\text{app}}}(v^2 - 2vv_0 + v_0^2) \\ &= \frac{1}{2}\frac{M}{F_{\text{app}}}(v^2 - v_0^2); \end{aligned} \tag{5}$$

thus

$$\boxed{\frac{1}{2}Mv^2 - \frac{1}{2}Mv_0^2 = F_{\text{app}} \times (x - x_0).} \tag{6}$$

If we define $\frac{1}{2}Mv^2$ as the *kinetic energy* of the particle, then the left-hand side of (6) is the change in kinetic energy. The change is caused by the force F_{app} acting for a distance $(x - x_0)$. It is obviously a good definition of *work* to call $F_{\text{app}}(x - x_0)$ the *work done by the applied force on the particle.* With these *definitions* (6) says that the work done by the applied force is equal to the change of kinetic energy of the particle. This is all a matter of definition, but the definitions are useful and they are consistent with Newton's second law. In talking about work, one must always specify "*work done by what.*"

If $M = 20$ gm and $v = 100$ cm/sec, the kinetic energy

$$K = \tfrac{1}{2}Mv^2 = \tfrac{1}{2}(20)(10^4) = 1 \times 10^5 \text{ gm-cm}^2/\text{sec}^2$$
$$= 1 \times 10^5 \text{ ergs.} \tag{7}$$

The *erg* is the unit of energy in the CGS system of units. If a 100-dyne force is applied through a distance of 10^3 cm,

$$F_{\text{app}}(x - x_0) = (10^2)(10^3) = 10^5 \text{ dyne-cm} = 10^5 \text{ ergs.} \tag{8}$$

One erg is the quantity of work performed by a force of one dyne acting for one centimeter. Work has the dimensions

$$[\text{Work}] \sim [\text{force}][\text{distance}] \sim [\text{mass}][\text{acceleration}][\text{distance}]$$
$$\sim \left[\mathrm{M}\frac{\mathrm{L}}{T^2}\mathrm{L}\right] \sim [\mathrm{ML}^2T^{-2}] \sim [\text{energy}].$$

In the MKS system the unit of work is the *joule*, which is the work done by a force of one newton acting through one meter. To convert joules to ergs, multiply by 10^7 the value of the work expressed in joules. To convert newtons to dynes, multiply by 10^5 the value of the force expressed in newtons.

Suppose the particle is not in intergalactic space but is at rest at a height h above the surface of the earth ($x_0 = h$; $v_0 = 0$). The gravitational force $F_G = -Mg$ pulls downward on the body. As the body falls toward the surface of the earth, the work done by gravity is equal to the gain in kinetic energy of the body:

$$W(\text{by gravity}) = F_G \times (x - x_0), \tag{9}$$

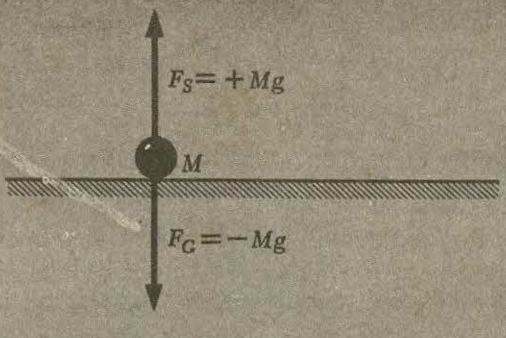
or, at the surface ($x = 0$) of the earth,

$$W(\text{by gravity}) = (-Mg)(0 - h) = Mgh = \tfrac{1}{2}Mv^2 - \tfrac{1}{2}Mv_0^2 = \tfrac{1}{2}Mv^2, \tag{10}$$

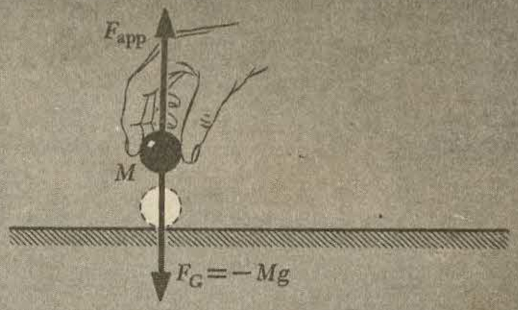
where v is the velocity of the body on reaching the surface of the earth and v_0 is its initial velocity. (In this particular example $v_0 = 0$.) Equation (10) suggests we may say that at height h the body has *potential energy* (capacity to do work or to gain kinetic energy) of Mgh relative to the earth's surface.

What happens to the potential energy when a particle at rest on the earth's surface is raised to a height h? To raise the body, we must apply an upward force $F_{\text{app}}(= -F_G)$ to the body. Now $x_0 = 0$; $x = h$. We do work

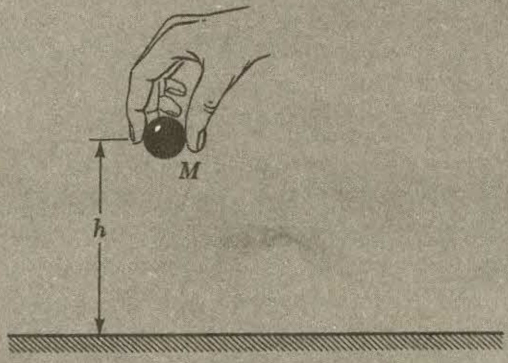
$$W(\text{by us}) = F_{\text{app}} \times (x - x_0) = (Mg)(h) = Mgh \tag{11}$$



A mass M at rest on the earth's surface experiences two equal and opposite forces: F_G, the attractive gravitational force; F_S, the force exerted on M by the supporting surface.



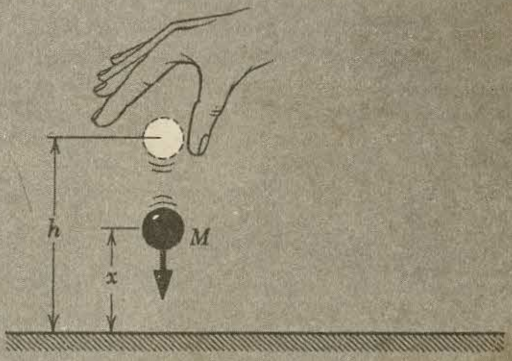
To raise M at constant velocity requires an applied force $F_{\text{app}} = +Mg$.



The amount of work done in raising M to height h is

$$W = F_{\text{app}} \times h = +Mgh.$$

The potential energy U of the mass M is thereby increased by an amount Mgh.

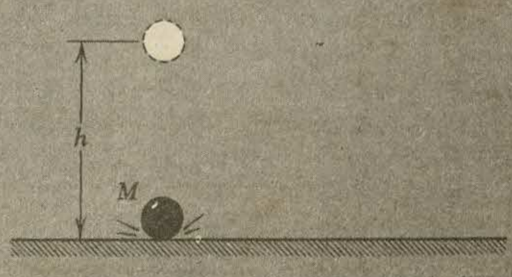


If the mass is released, the potential energy U decreases and the kinetic energy K increases, but their sum remains constant. At height x,

$$U(x) = Mgx,$$

and

$$K(x) = \tfrac{1}{2}Mv(x)^2 = Mg(h - x).$$



Just before the mass strikes the ground, all the potential energy the mass had at h has been converted to kinetic energy.

on the body, thereby giving the body the potential energy Mgh which, as we have said earlier, it has at height h.

Note that the work done by gravity (10) on the falling body is equal to the work we do (11) against gravity in lifting the particle up.

The units $[F][L] = [M][L^2]/[T^2]$ of potential energy are the same as those of kinetic energy. If $F_{\text{app}} = 10^3$ dynes and $h = 10^2$ cm, the potential energy is $(10^3)(10^2) = 10^5$ dyne-cm $= 10^5$ ergs. We denote the potential energy by U or by P.E.

In (10) we see that the potential energy Mgh can be converted entirely into kinetic energy since the work done by the applied force F_{app} in building up the potential energy is the same as the kinetic energy developed in (10) and in (6). We must always specify the point to which the potential energy is referred. Potential energy is introduced as a matter of bookkeeping; we have shown that our bookkeeping in this example does not lead to contradictions. Changes in potential energy are physically meaningful, for the change tells us how much kinetic energy is developed.

If in (10) we let v denote, not the velocity after falling a distance h, but the velocity after falling a distance $(h - x)$, then the equation analogous to (10) is

$$\tfrac{1}{2}Mv^2 = Mg(h - x), \tag{12}$$

or

$$\tfrac{1}{2}Mv^2 + Mgx = Mgh = E, \tag{13}$$

where E is a constant having the value Mgh. Because E is a constant we have in (13) a statement of the *law of conservation of energy:*

$$\begin{aligned} E &= K + U \\ &= \text{kinetic energy} + \text{potential energy} = \text{const} = \text{total energy}. \end{aligned}$$

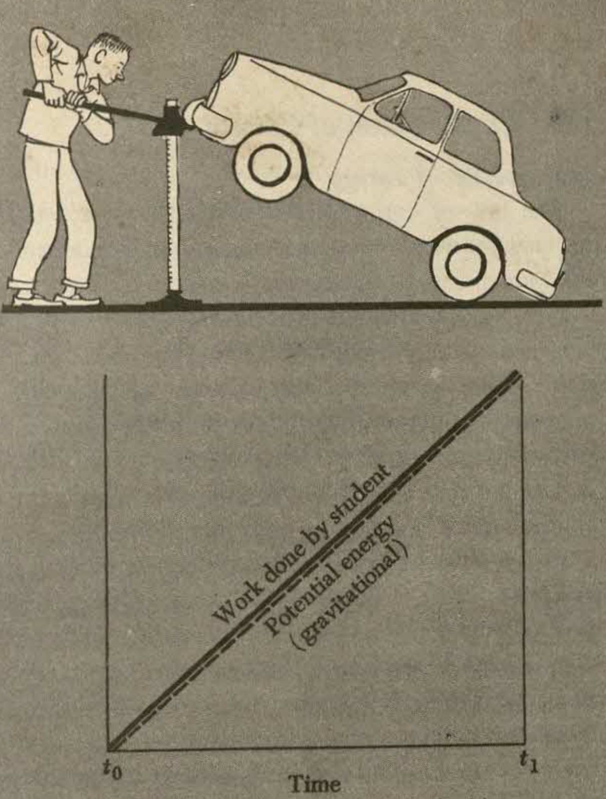
In (13) the term Mgx is the potential energy, where we have chosen $x = 0$ as the zero of potential energy. The symbol E denotes the total energy, which is constant with time for an isolated system.

Sometimes it is convenient to call $E = K + U$, the sum of kinetic and potential energy contributions, the *energy function.* The kinetic energy contribution K is equal to $\tfrac{1}{2}Mv^2$. The potential energy depends on the force acting. The potential energy U has the essential property that $-\int F\,dx = U$,

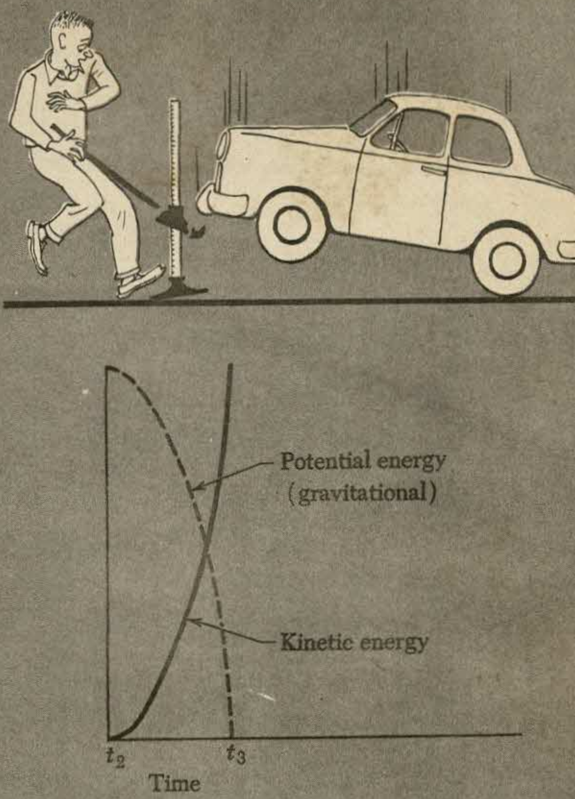
or

$$F = -\frac{dU}{dx}, \tag{14}$$

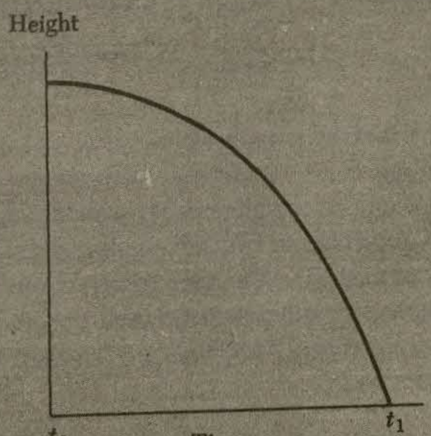
where F, the force acting on the particle, results from intrinsic interactions, such as electrical or gravitational interactions. (In the above example, $U = Mgx$, so that $F = F_G = -Mg$.) We now discuss these ideas in greater detail and generality.



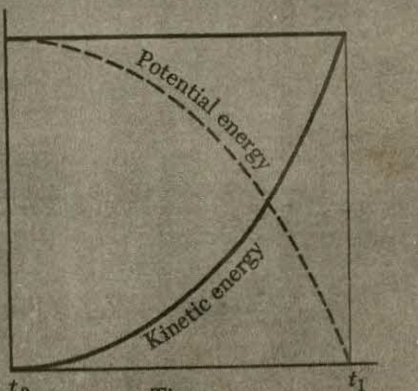
Work done by student vs time in jacking a car to change a tire. (The work done on a 1000-kg small car in raising its center of mass by 10 cm will be $Fh = Mgh \cong (10^3 \times 10^3 \text{ gm}) \times (10^3 \text{ cm/sec}^2)(10 \text{ cm}) = 10^{10}$ ergs.) The work done appears as gravitational potential energy.



Here the jack has slipped and the car falls back. The potential energy is converted into kinetic energy. After the car makes contact with the ground, the kinetic energy is converted to heat in the shock absorbers, springs, and tires.



Height vs time for a body falling toward the earth, starting at rest.



Potential energy of the falling body vs time, and kinetic energy of the falling body vs time. The total energy, which is constant, is the sum of the kinetic, plus the potential, energy.

Conservation of Energy

The law of conservation of energy states that for a system of particles with interactions not explicitly† dependent on the time the *total energy of the system is constant.* We accept this result as a very well established experimental fact. More specifically, the law tells us there exists some scalar function [such as the function $\frac{1}{2}Mv^2 + Mgx$ in (13)] of the positions and velocity of the constituent particles which is invariant with respect to a change in time, provided there is no explicit change in the external interaction during the time interval considered. For example, the elementary charge e must not change with time. Besides the energy function, there are other functions which are constant in the conditions specified here. (We treat other functions in Chap. 6, under conservation of linear and angular momentum.) The energy is a scalar constant of the motion. We interpret the phrase *external interaction* to include any change in the laws of physics or in the values of the fundamental physical constants (such as g or e or m) during the relevant time interval. Remember that the law gives us no new information not contained in the equations of motion $\mathbf{F} = M\mathbf{a}$.

The central problem is to find an expression for the energy function which has the desired time invariance and which is consistent with $\mathbf{F} = M\mathbf{a}$. By consistent we mean that, for example,

$$\frac{d}{dx}E \equiv \frac{d}{dx}(K + U) = \frac{dK}{dx} - F_x = 0$$

is identical with $F_x = Ma_x$. You can check this for (13), to find

$$Mg = Mv\frac{dv}{dx} \equiv M\frac{dx}{dt}\frac{dv}{dx} \equiv M\frac{dv}{dt}.$$

This is the fundamental problem of classical mechanics, and its formal solution can be given in many ways, some quite elegant. The Hamiltonian formulation of mechanics, in particular, is very well suited to reinterpretation in the language of quantum mechanics. But here at the beginning of our course we need a simple direct formulation more than we need the generality of the Hamiltonian or Lagrangian formulations, which are usually the subject of later courses.‡

---

† Consider the system with the particles permanently frozen in place; then a force which depends on time is said to depend *explicitly* on time.

‡ The derivation of the Lagrangian equations of motion requires several elementary results of the calculus of variations; this problem deterred us from going further here.

Work. We define the work W done on a particle by a constant applied force $\mathbf{F}_{\text{app}}$ in a displacement $\Delta\mathbf{r}$ as

$$W = \mathbf{F}_{\text{app}} \cdot \Delta\mathbf{r} = F\,\Delta r \cos{(\mathbf{F}_{\text{app}}, \Delta\mathbf{r})}, \tag{15}$$

in conformity with the definition which followed (6) above. In the two paragraphs which follow we write $\mathbf{f}$ for $\mathbf{F}_{\text{app}}$ to avoid a clumsy notation.

Suppose $\mathbf{f}$ is not constant, but is a function $\mathbf{f}(\mathbf{r})$ of the position $\mathbf{r}$. If the path may be decomposed into N line segments within each of which $\mathbf{f}(\mathbf{r})$ is constant, then we may write

$$\begin{aligned} W &= \mathbf{f}(\mathbf{r}_1) \cdot \Delta\mathbf{r}_1 + \mathbf{f}(\mathbf{r}_2) \cdot \Delta\mathbf{r}_2 + \cdots + \mathbf{f}(\mathbf{r}_N) \cdot \Delta\mathbf{r}_N \\ &\equiv \sum_{j=1}^{N} \mathbf{f}(\mathbf{r}_j) \cdot \Delta\mathbf{r}_j, \end{aligned} \tag{16}$$

where the symbol Σ stands for the sum indicated. Equation (16) is strictly valid only in the limit of infinitesimal displacements $d\mathbf{r}$, because in general a curved path cannot be decomposed exactly into a finite number of line segments.

The limit

$$\lim_{\Delta\mathbf{r}\to 0} \sum_j \mathbf{f}(\mathbf{r}_j) \cdot \Delta\mathbf{r}_j = \int_{\mathbf{r}_A}^{\mathbf{r}_B} \mathbf{f}(\mathbf{r}) \cdot d\mathbf{r} \tag{17}$$

is the integral of the projection of $\mathbf{f}(\mathbf{r})$ on the displacement vector $d\mathbf{r}$. The integral is called the *line integral* of $\mathbf{f}$ from A to B. The work done in the displacement by the applied force is defined as

$$\boxed{W(A \to B) \equiv \int_A^B \mathbf{F}_{\text{app}} \cdot d\mathbf{r}.} \tag{18}$$

Kinetic Energy. We now return to the free particle in intergalactic space. We want to generalize (6),

$$\tfrac{1}{2}Mv^2 - \tfrac{1}{2}Mv_0^2 = F_{\text{app}} \times (x - x_0), \tag{6}$$

to include applied forces which vary in direction and magnitude. By substituting $\mathbf{F}_{\text{app}} = M\dot{\mathbf{v}}$ into (18) we find for the work done by the applied force

$$\boxed{W(A \to B) = M \int_A^B \frac{d\mathbf{v}}{dt} \cdot d\mathbf{r}.} \tag{19}$$

Now

$$d\mathbf{r} = \frac{d\mathbf{r}}{dt}\,dt = \mathbf{v}\,dt, \tag{20}$$

so that

$$W(A \to B) = M\int_A^B \left(\frac{d\mathbf{v}}{dt}\cdot\mathbf{v}\right) dt. \tag{21}$$

But we can rearrange the integrand:

$$\frac{d}{dt}v^2 = \frac{d}{dt}(\mathbf{v}\cdot\mathbf{v}) = 2\frac{d\mathbf{v}}{dt}\cdot\mathbf{v}, \tag{22}$$

so that

$$2\int_A^B \left(\frac{d\mathbf{v}}{dt}\cdot\mathbf{v}\right) dt = \int_A^B \left(\frac{d}{dt}v^2\right) dt = \int_A^B d(v^2) = v_B{}^2 - v_A{}^2. \tag{23}$$

On substitution in (21) we have an important result:

$$\boxed{W(A \to B) = \int_A^B \mathbf{F}_{\text{app}}\cdot d\mathbf{r} = \frac{1}{2}Mv_B{}^2 - \frac{1}{2}Mv_A{}^2,} \tag{24}$$

for the free particle. This is a generalization of (6).

We call

$$K \equiv \tfrac{1}{2}Mv^2 \tag{25}$$

the *kinetic energy* and denote it by the symbol K. We see from (24) that our definitions of work and kinetic energy have the property that *the work done on a free particle by an arbitrary applied force is equal to the change in the kinetic energy of the particle:*

$$W(A \to B) = K_B - K_A. \tag{25a}$$

Example. Free fall. (a) If the x direction is normal to the surface of the earth and directed upward, the gravitational force is $\mathbf{F}_G = -Mg\,\hat{\mathbf{x}}$, where g is the acceleration of gravity and has the approximate value 980 cm/sec². Calculate the work done by gravity when a mass of 100 gm falls through 10 cm.

Here we can set

$$\mathbf{r}_A = 0; \qquad \mathbf{r}_B = -10\hat{\mathbf{x}}; \qquad \Delta\mathbf{r} = \mathbf{r}_B - \mathbf{r}_A = -10\hat{\mathbf{x}}.$$

From (15), the work done by gravity is

$$\begin{aligned} W = \mathbf{F}_G\cdot\Delta\mathbf{r} &= (-Mg\hat{\mathbf{x}})\cdot(-10\hat{\mathbf{x}}) \\ &\approx (10^2)(10^3)(10)\,\hat{\mathbf{x}}\cdot\hat{\mathbf{x}} \approx 10^6 \text{ ergs}. \end{aligned}$$

Here we have let the gravitational force play the role of the force $\mathbf{F}_{\text{app}}$.

(*b*) If the particle in (*a*) was initially at rest, what is its kinetic energy and its velocity at the end of its 10-cm fall?

The initial value K_A of the kinetic energy is zero; the terminal value K_B according to (25*a*) is equal to the work done by gravity on the particle, so that

$$K_B = \tfrac{1}{2}Mv_B^2 \simeq 10^6 \text{ ergs}, \tag{26}$$

whence $v_B^2 \simeq 2(10^6 \text{ ergs})/(100 \text{ gm}) \simeq 2 \times 10^4 \text{ cm}^2/\text{sec}^2$. This result agrees with what we would calculate from the elementary relation $v^2 = 2gh$ for constant acceleration from rest, for $v^2 \simeq 2(10^3 \text{ cm/sec}^2)(10 \text{ cm}) \simeq 2 \times 10^4 \text{ cm}^2/\text{sec}^2$.

This is an example of what we mean when we say that the results obtained from the conservation law must be consistent with the equations of motion. Here using conservation of energy, we obtained the same result as by employing an equation $v^2 = 2gh$ derived from the equation of motion $\mathbf{F} = M\mathbf{a}$.

Example. Linear restoring force. A particle is subject to a linear restoring force in the x direction. A linear restoring force is one which is directly proportional to the displacement measured from some fixed point and in a direction tending to reduce the displacement. If we take the fixed point as the origin,

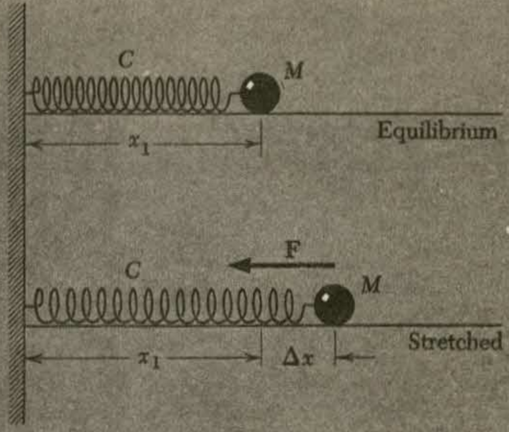
$$\mathbf{F} = -Cx\hat{\mathbf{x}}, \qquad \text{or} \qquad F_x = -Cx, \tag{27}$$

where C is some positive constant, the *force constant.* This is called Hooke's law. For sufficiently small displacements such a force may be produced by a stretched or compressed spring. For large elastic displacements we must add terms in higher powers of x to (27), as we do in Chap. 7. The sign of the force is such that the particle is always attracted toward the origin $x = 0$.

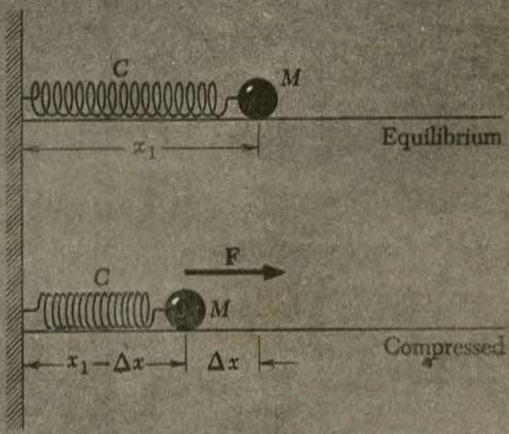
(*a*) With the particle attached to the spring, we now apply an external force which takes the particle from a point x_1 to another point x_2. What is the work done by the applied force on the particle in the displacement?

Here the force on the particle is a function of the position. To calculate the work done by the applied force we use the definition (18), with $\mathbf{F}_{\text{app}} = -\mathbf{F} = +Cx\hat{\mathbf{x}}$:

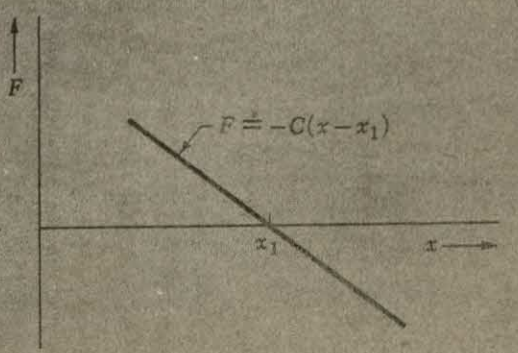
$$W(x_1 \to x_2) = \int_{x_1}^{x_2} \mathbf{F}_{\text{app}} \cdot d\mathbf{r} = C\int_{x_1}^{x_2} x\,dx = \frac{1}{2}C(x_2^2 - x_1^2). \tag{28}$$



A massless spring is connected to mass M. If the spring is stretched a small amount Δx, it exerts a restoring force $F = -C\,\Delta x$ on M in the direction shown. Here C denotes the force constant of the spring.



If the spring is compressed an amount $-\Delta x$, the spring exerts a restoring force on M equal to $-C(-\Delta x) = C\,\Delta x$, as shown.



The restoring force for small displacements from x_1 is proportional to the displacement.

(Here $\mathbf{F}_{\text{app}}$ is the force applied *by someone* against the force of the spring, and it is *not* the force which the spring itself applies to the mass. The mass gains no kinetic energy in the process if the two forces are equal and opposite.) If the initial point x_1 is taken as the origin, then $x_1 = 0$ and (28) reduces to

$$W(0 \rightarrow x) = \tfrac{1}{2}Cx^2. \tag{29}$$

This is a famous result: It says that the work done on the system by the applied force is proportional to the square of the displacement.

(*b*) If the particle of mass M is released at rest at the position x_{max}, what is its kinetic energy when it reaches the origin?

We obtain the answer directly from (24): the work done by the *spring* in going from x_{max} to the origin is

$$W(x_{\text{max}} \rightarrow 0) = \tfrac{1}{2}Mv_1^2, \tag{30}$$

where we have used the fact that $v = 0$ at x_{max}; the particle is assumed to be at rest there. The velocity at the origin is v_1. Thus

$$\tfrac{1}{2}Cx_{\text{max}}^2 = \tfrac{1}{2}Mv_1^2 \tag{31}$$

is the kinetic energy at the origin $x = 0$.

(*c*) What is the connection between the velocity of the particle at the origin and the maximum displacement x_{max}?

From (31),

$$v_1^2 = \frac{C}{M}x_{\text{max}}^2, \tag{32}$$

or

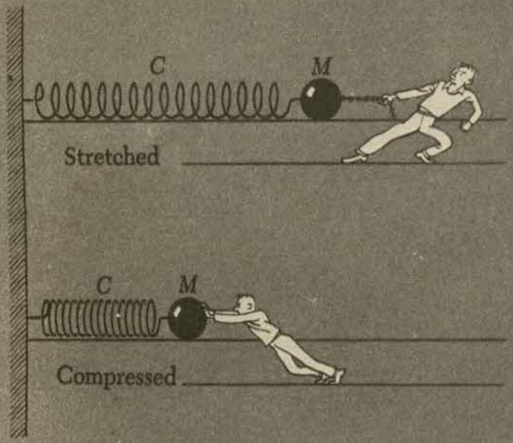
$$v_1 = \pm\sqrt{\frac{C}{M}}\,x_{\text{max}}. \tag{33}$$

Power. The power P is the time rate of transfer of energy. We defined the work done on the particle in a displacement $\Delta\mathbf{r}$ by an applied force as

$$\Delta W = \mathbf{F}_{\text{app}} \cdot \Delta\mathbf{r}. \tag{34}$$

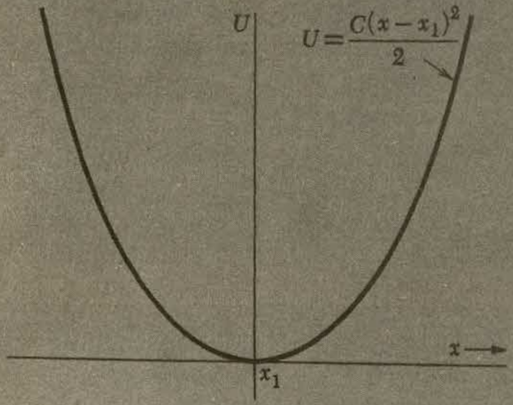
The rate at which work is done by the force is

$$\frac{\Delta W}{\Delta t} = \mathbf{F}_{\text{app}} \cdot \frac{\Delta\mathbf{r}}{\Delta t}. \tag{35}$$



In order to stretch (or compress) the spring, you must exert a force in opposition to the restoring force. In displacing the spring an amount Δx from the equilibrium position x_1, you do work

$$W = \int_{x_1}^{x_1+\Delta x} C(x - x_1)\,dx = \tfrac{1}{2}C(\Delta x)^2.$$



In doing this work you increase the potential energy of the spring-mass system. A spring-mass system displaced $\Delta x = x - x_1$ from equilibrium has potential energy

$$U = \tfrac{1}{2}C(\Delta x)^2 = \tfrac{1}{2}C(x - x_1)^2.$$

In the limit $\Delta t \rightarrow 0$ we have the power

$$\boxed{P = \frac{dW}{dt} = \mathbf{F}_{\text{app}} \cdot \frac{d\mathbf{r}}{dt} = \mathbf{F}_{\text{app}} \cdot \mathbf{v}.} \tag{36}$$

From the power $P(t)$ as a function of time we can write the work input as

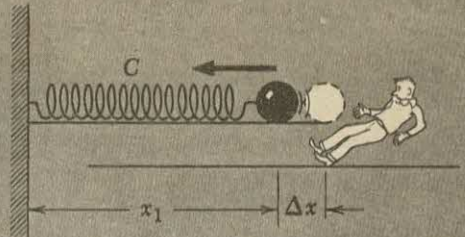
$$W(t_1 \rightarrow t_2) = \int_{t_1}^{t_2} P(t)\, dt.$$

In the CGS system *the unit of power is one erg per second.* In the MKS system the unit of power is one joule per second, which is called one *watt.* To find the power in erg/sec, multiply the power expressed in watts by 10^7. To obtain the power in watts from the value expressed in horsepower, multiply by 746, approximately.

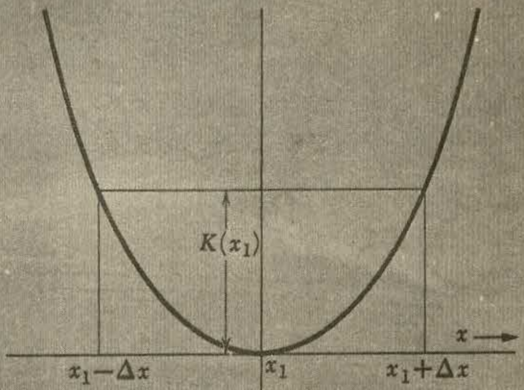
Conservative Forces. A force is conservative if the work $W(A \rightarrow B)$ done by the force in moving the particle from A to B is independent of the path by which the particle is moved between A and B. This is illustrated by Eq. (24). Thus, because $W(A \rightarrow B) = -W(B \rightarrow A)$, we see that if the particle is moved once around a closed path the work done by a conservative force is zero.

We can easily see that a central force is conservative. A *central force* exerted by one particle on another is a force whose magnitude depends only on the separation of the particles and whose direction lies along the line joining the particles. In the figure opposite, a central force is directed away from (or toward) the center at the point O. Two paths, labeled 1 and 2, connect points A and B as shown. The dashed curves are sectors of circles centered at O. Consider the quantities $(\mathbf{F}_1 \cdot d\mathbf{r}_1)$ and $(\mathbf{F}_2 \cdot d\mathbf{r}_2)$ evaluated on the path segments lying between the dashed circles. (We may regard $\mathbf{F} \cdot d\mathbf{r} = F\,dr\cos\theta$ equally well as the projection of $\mathbf{F}$ on $d\mathbf{r}$ or $d\mathbf{r}$ on $\mathbf{F}$.) Now the magnitudes F_1 and F_2 are equal on the two segments because they lie at equal distances from the point O; the projections $dr\cos\theta$ of the path segments on the respective vectors $\mathbf{F}$ are equal, because the separation of the circles measured along the direction of $\mathbf{F}_1$ is equal to the separation measured along $\mathbf{F}_2$. Therefore

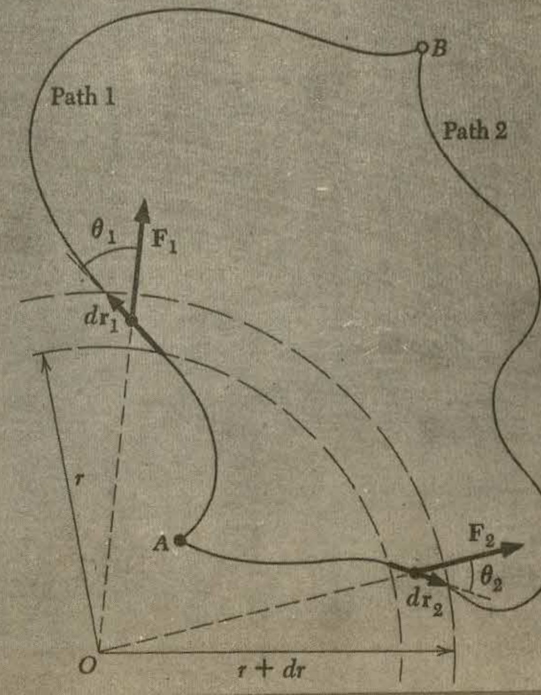
$$(\mathbf{F}_1 \cdot d\mathbf{r}_1) = (\mathbf{F}_2 \cdot d\mathbf{r}_2) \tag{37}$$



If the spring-mass system is stretched Δx and then released, U will initially decrease and K will increase.



At $x = x_1$, $U = 0$ and $K(x_1) = \frac{1}{2}C(\Delta x)^2$, as shown.



on the path segments considered. But the identical argument can be employed repeatedly for every comparable path segment, so that

$$\underset{\text{(Path 1)}}{\int_A^B \mathbf{F} \cdot d\mathbf{r}} = \underset{\text{(Path 2)}}{\int_A^B \mathbf{F} \cdot d\mathbf{r}}. \tag{38}$$

Forces with the property that

$$W(A \rightarrow B) = \int_A^B \mathbf{F} \cdot d\mathbf{r} \tag{39}$$

is independent of the path are called *conservative* forces. For conservative forces the work done around a closed path is *zero.*

Suppose the force depends on the velocity with which the path is traversed. (The force on a charged particle in a magnetic field depends on velocity.) Can such a force be conservative? It turns out that the important fundamental velocity-dependent forces are conservative, because their direction is *perpendicular* to the direction of motion of the particle, so that $\mathbf{F} \cdot d\mathbf{r}$ is zero. You can see this for the Lorentz force (Chap. 4), which is proportional to $\mathbf{v} \times \mathbf{B}$. Frictional forces are not really fundamental forces, but they are velocity-dependent and are not conservative.

All of our discussion presupposes *two-body forces.* This is an important assumption; it is likely that students of this course will be called upon in their research careers to do battle with many-body forces. A discussion of what is involved in the two-body assumption is given in Vol. II.

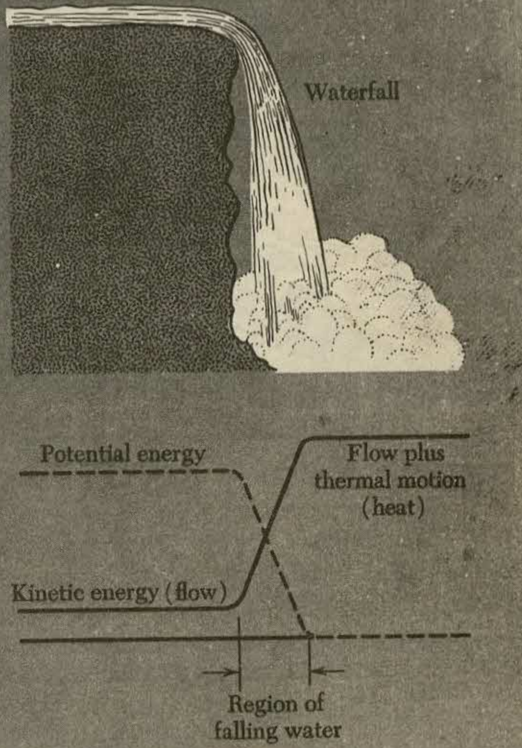
It is known experimentally that $W(A \rightarrow B)$ is independent of the path for gravitational forces and for electrostatic forces. This result for interactions between elementary particles is inferred from scattering experiments; for gravitational forces the result is inferred from the accuracy of the prediction of planetary and lunar motions, as discussed in the Historical Notes. We also know that the earth has made about 4×10^9 complete orbits around the sun without any important change in distance to the sun, as judged from geologic evidence on the surface temperature of the earth. The relevant geologic evidence extends back perhaps 10^9 years and cannot be taken as entirely conclusive because of the numerous factors, including solar output, which affect the temperature, but the observation is suggestive. Further examples are discussed in the Historical Notes at the end of the chapter.

We need to say more about central vs noncentral forces. In consideration of the force between two particles there are two possibilities: (1) The particles have no coordinates other than their positions. (2) One or both particles have a physically distinguished axis. In the first possibility there can *only* be a central force, while in the second the specification that the particle be moved from A to B is incomplete—we have also to specify that the axis be kept in the same direction relative to something. A bar magnet has a physically distinguished axis: If we move the magnet bodily around a closed path in a uniform magnetic field, we may or may not do a net amount of work on the magnet. If the magnet ends up at the same location and in the *same orientation* as it started out, no work is done. If the location is the same, but the orientation is different, work will have been done. (The work may have a positive or negative sign.)

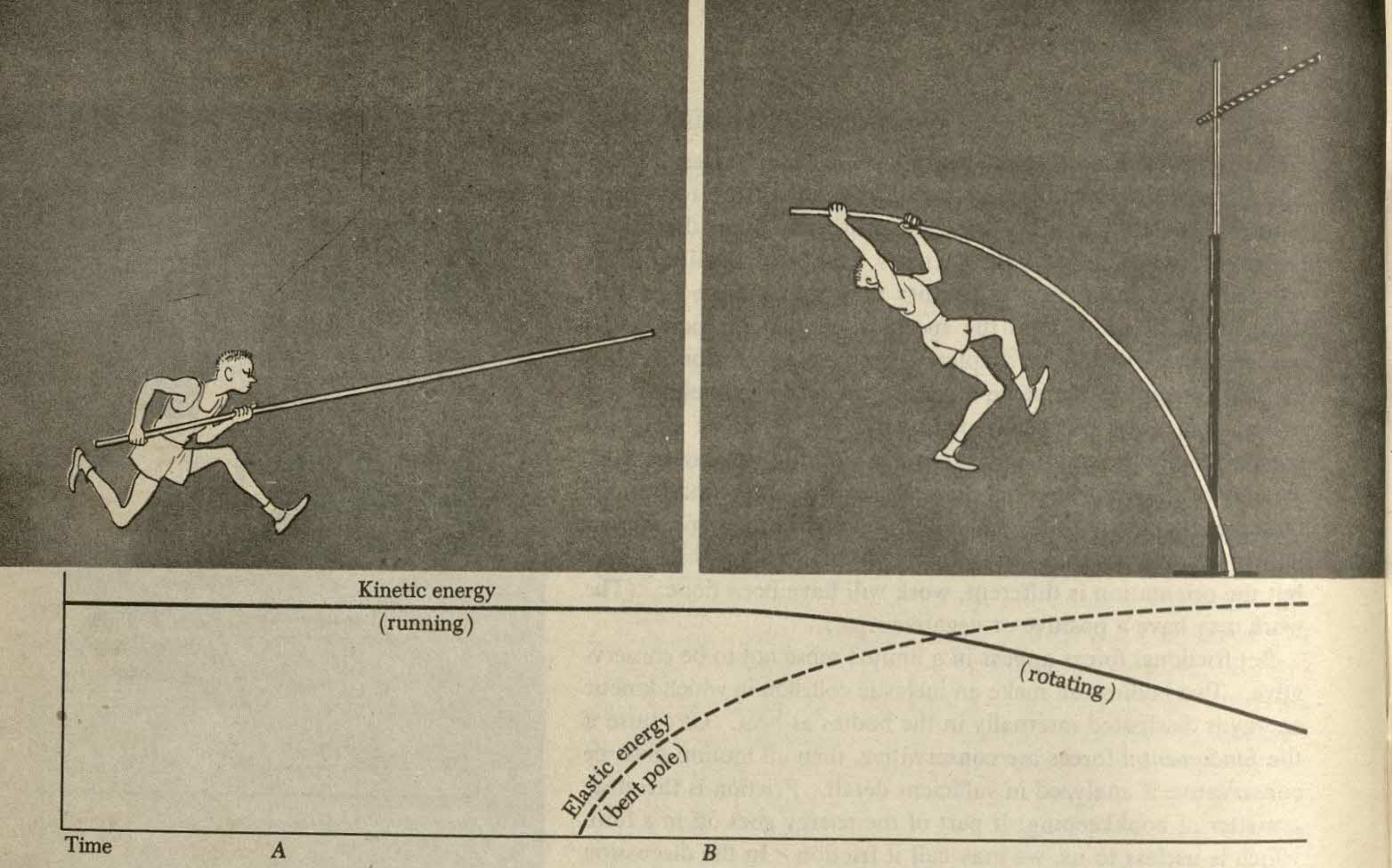
But frictional forces appear in a limited sense not to be conservative. Two bodies can make an inelastic collision in which kinetic energy is dissipated internally in the bodies as heat. Of course if the *fundamental* forces are conservative, then all motion must be conservative if analyzed in sufficient detail. Friction is therefore a matter of bookkeeping: If part of the energy goes off in a form which is useless to us, we may call it friction. In the discussion of conservation of momentum in Chap. 3 we considered an inelastic collision of two particles. The kinetic energy was not conserved; but the sum of the kinetic energy and the internal excitation energy for the two particles was called the total energy and was assumed to be conserved, in agreement with all known experiments.

Potential Energy. We have seen that with a suitable applied force $\mathbf{F}_{\text{app}}$ we can just counterbalance any other forces (such as gravitational) which act on a particle, and we can thereby move the particle very slowly from one position to another, without developing kinetic energy in the process. We have said that in such a situation the work done in moving the particle will appear as potential energy. We define the difference in the potential energy of the particle at points **B** and **A** as equal to the work done by such an applied force on the particle in moving it to **B** from **A**:

$$\boxed{U(\mathbf{B}) - U(\mathbf{A}) = W(\mathbf{A} \to \mathbf{B}) = \int_A^B \mathbf{F}_{\text{app}} \cdot dr.} \tag{40}$$



The water at the top of the waterfall has gravitational potential energy, which in falling is converted into kinetic energy. A mass M of water in falling from a height h loses potential energy Mgh and gains kinetic energy $\frac{1}{2}M(v^2 - v_0^2) = Mgh$. (The velocity v is determined by this equation if the initial velocity v_0 of the water is known.) The kinetic energy of the falling water can be converted in a powerhouse into the rotational kinetic energy of a turbine; otherwise the kinetic energy of the falling water is converted at the foot of the falls into thermal energy or heat. Thermal energy is simply the random energy of molecular motion in the water. (At a high temperature the random molecular motion is more vigorous than at a low temperature.)



Motion of a pole vaulter. At *A* his energy is all kinetic, associated with his running velocity. At *B* he puts the forward end of the pole on the ground and (especially with the new fiberglass poles) stores elastic potential energy in the pole by bending it. At *C* he is rising in the air; he has considerable kinetic energy left, now associated with his rotational velocity about the lower end of the pole. He has potential energy

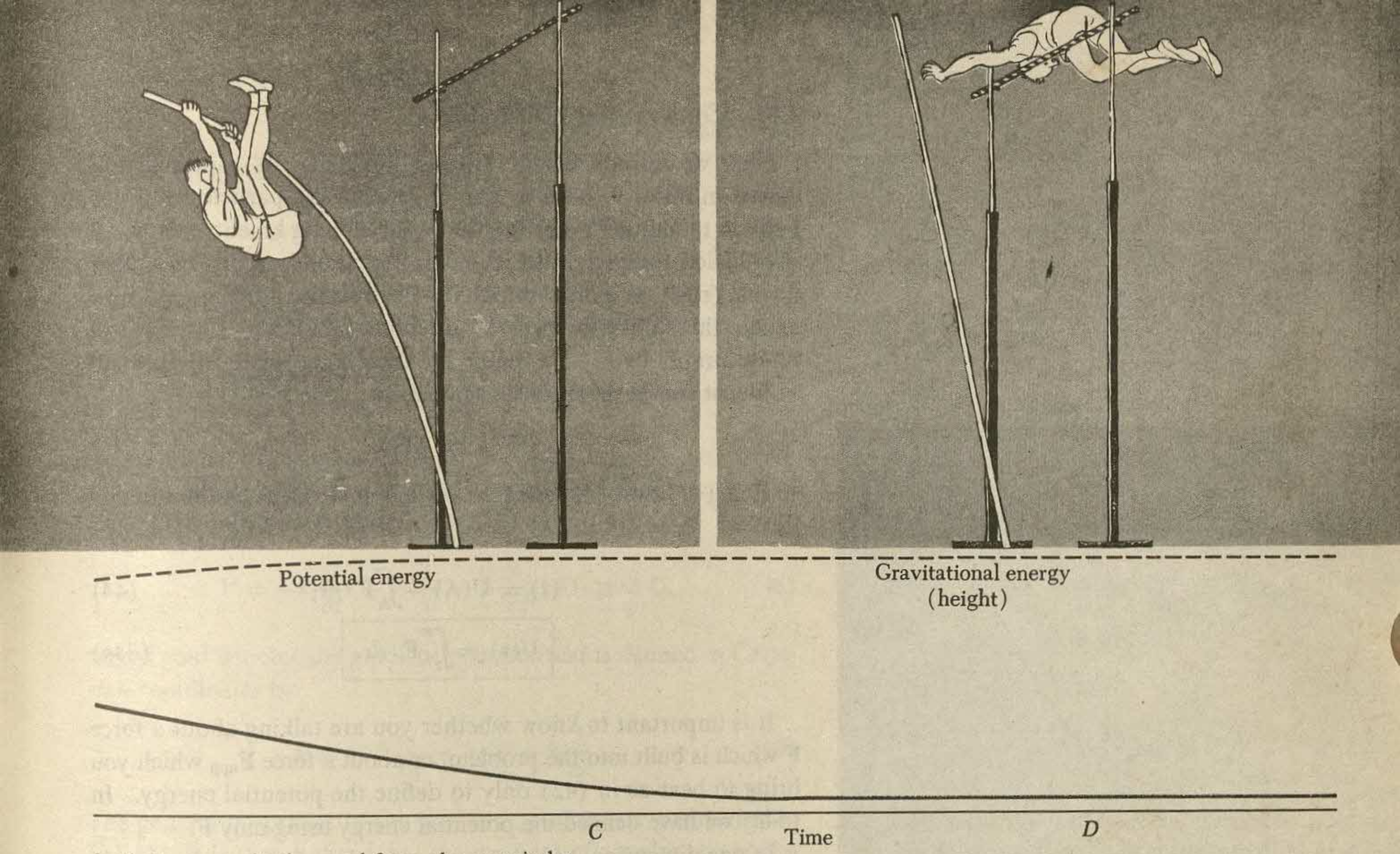
We assume here that the forces are conservative; then $U(\mathbf{r})$ will be a single-valued scalar function† of position, and $U(\mathbf{B}) - U(\mathbf{A})$ will be equal to the increase in kinetic energy of the particle on moving back from **B** to **A** *if the applied force is removed.*

By fixing the value of U at some point, say **A**, the relation (40) defines $U(\mathbf{r})$ at any other point **r**, for then

$$\boxed{U(\mathbf{r}) = U(\mathbf{A}) + \int_{\mathbf{A}}^{\mathbf{r}} \mathbf{F}_{\text{app}} \cdot d\mathbf{r}.} \qquad (41)$$

---

† A function $f(x)$ is single valued if to each value of x there corresponds one and only one value of $f(x)$. For example, $\sin x$ is a single-valued function of x, but $\sin^{-1} x$ is not.



both from gravitation and from the remainder of the elastic energy of the pole. At *D*, going over the bar, his kinetic energy is low because he is moving slowly; his potential energy (gravitational) is high. The total energy is not always constant in the pole vault both because of friction (external and muscular) and because, while bending the pole, the vaulter is doing work.

The value of the constant $U(\mathbf{A})$ is not defined, so that (41) defines $U(\mathbf{r})$ only within a constant, which we may select arbitrarily at our convenience. Only the difference $U(\mathbf{r}) - U(\mathbf{A})$ of the potential energy between two points **r** and **A** ever has a physical meaning. In many problems **A** is taken at infinite distance and $U(\infty)$ is defined to be zero. Then

$$\boxed{U(\mathbf{r}) = \int_{\infty}^{\mathbf{r}} \mathbf{F}_{\text{app}} \cdot d\mathbf{r} = W(\infty \to \mathbf{r});} \qquad (42)$$

with this convention the potential energy at **r** *is equal to the work done by an applied force in moving the particle from infinity to* **r**.

Here we introduced the applied force $\mathbf{F}_{\text{app}}$ as a pedagogical device in order to have a way of moving the particle from one point **A** to another point **B** without developing kinetic energy. If the particle interacts with its environment and experiences other forces (such as gravitational or electrostatic forces), we may denote these intrinsic forces or interaction forces by $\mathbf{F}_{\text{int}}$ or, most often, simply by **F**. We define the force $\mathbf{F}_{\text{app}}$ which we apply so as to just cancel the intrinsic force **F** by

$$\mathbf{F}_{\text{app}} = -\mathbf{F}_{\text{int}}, \tag{43}$$

so that (41) and (42) may be rewritten in terms of the interactions as

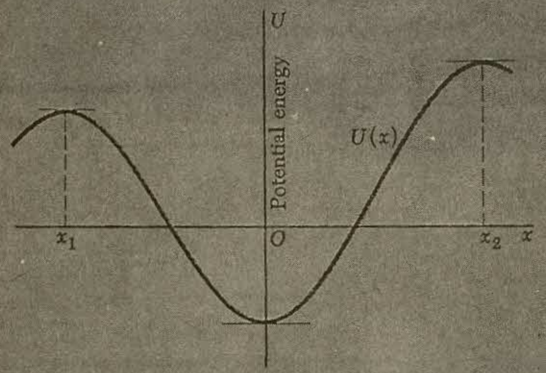
$$U(\mathbf{r}) = U(\mathbf{A}) - \int_{\mathbf{A}}^{\mathbf{r}} \mathbf{F} \cdot d\mathbf{r}; \tag{44}$$

$$\boxed{U(\mathbf{r}) = \int_{\mathbf{r}}^{\infty} \mathbf{F} \cdot d\mathbf{r}.} \tag{44a}$$

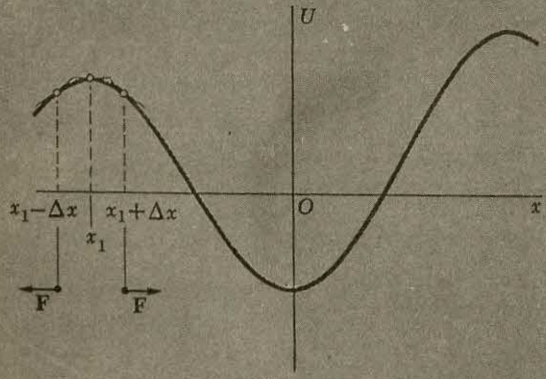
It is important to know whether you are talking about a force **F** which is built into the problem, or about a force $\mathbf{F}_{\text{app}}$ which you bring to bear as in (42) only to define the potential energy. In (44*a*) we have defined the potential energy using only **F**.

In one dimension

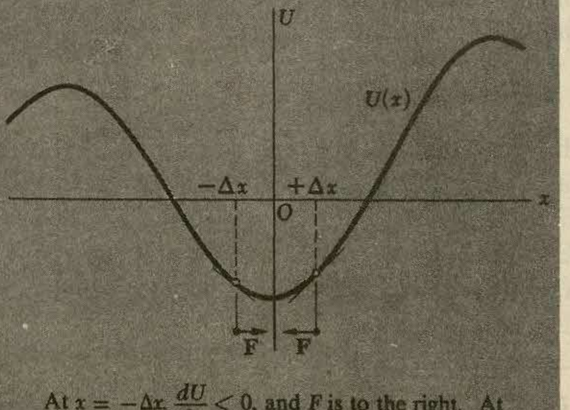
$$U(x) - U(A) = -\int_A^x F\,dx, \tag{45}$$



A one-dimensional potential energy function $U(x)$ plotted versus x. At points $x = x_1$, O, and x_2 we have $dU/dx = 0$ and thus the force F is zero at these points. These are, therefore, positions of equilibrium, not necessarily stable.



At point $x_1 - \Delta x$, $\frac{dU}{dx} > 0$, so that $F < 0$ (to the left). At point $x_1 + \Delta x$, $\frac{dU}{dx} < 0$, so $F > 0$ (to the right). A small displacement from x_1, therefore, results in a force tending to increase the displacement, and so x_1 is a position of *unstable* equilibrium.



At $x = -\Delta x$, $\frac{dU}{dx} < 0$, and F is to the right. At $x = +\Delta x$, $\frac{dU}{dx} > 0$, so that F is to the left. Thus $x = 0$ is a position of *stable* equilibrium. What about x_2?

whence on differentiation we have

$$\boxed{\frac{dU}{dx} = -F.} \tag{46}$$

This result may be checked by substitution of (46) into (45):

$$-\int_A^x F\,dx = \int_A^x \frac{dU}{dx}\,dx = \int_A^x dU = U(x) - U(A). \tag{47}$$

Equation (46) is an example of the general result that force is the negative of the space rate of change of potential energy. In three dimensions the expression analogous to (46) is†

$$\mathbf{F} = -\hat{\mathbf{x}}\frac{\partial U}{\partial x} - \hat{\mathbf{y}}\frac{\partial U}{\partial y} - \hat{\mathbf{z}}\frac{\partial U}{\partial z} \equiv -\text{grad } U, \tag{48}$$

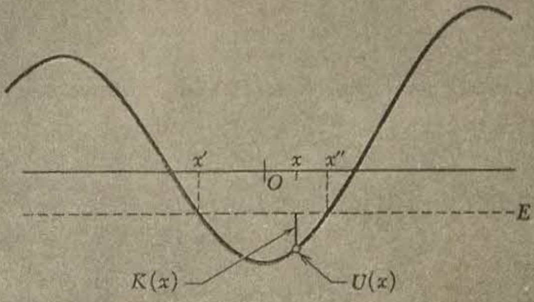
where grad denotes the gradient operator and is defined in Cartesian coordinates by

$$\text{grad} \equiv \hat{\mathbf{x}}\frac{\partial}{\partial x} + \hat{\mathbf{y}}\frac{\partial}{\partial y} + \hat{\mathbf{z}}\frac{\partial}{\partial z}. \tag{49}$$

The general properties of the gradient operator are considered in Vol. II. It is shown there that the gradient of a scalar is a vector whose direction is that of the maximum spatial rate of increase of the scalar, and whose magnitude is equal to the rate of change. The gradient of a scalar U is written variously as grad U or ∇U or $\partial U/\partial \mathbf{r}$. The operator ∇ is read as "del," and ∇U is read as "del you."

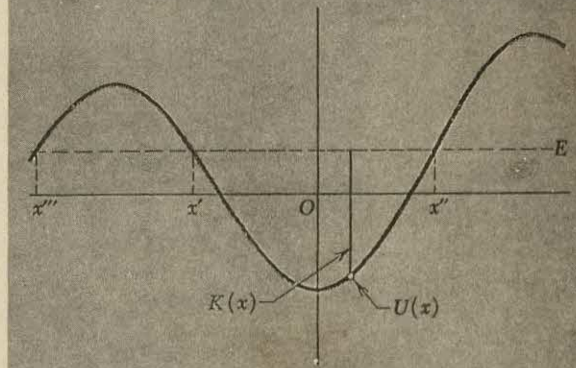
Potential Energy in an Electric Field. Suppose that the electric field intensity $\mathbf{E}(\mathbf{r})$ is known at every point of space. Suppose further that $\mathbf{E}(\mathbf{r})$ is produced by a fixed rigid distribution of electric charge. We saw in Chap. 3 that $\mathbf{E}(\mathbf{r})$ is defined as the force on a unit positive charge at rest, and according to Coulomb's law we can calculate $\mathbf{E}(\mathbf{r})$ as

$$\mathbf{E}(\mathbf{r}) = \sum_i \frac{q_i}{|\mathbf{r} - \mathbf{r}_i|^3}(\mathbf{r} - \mathbf{r}_i), \tag{50}$$

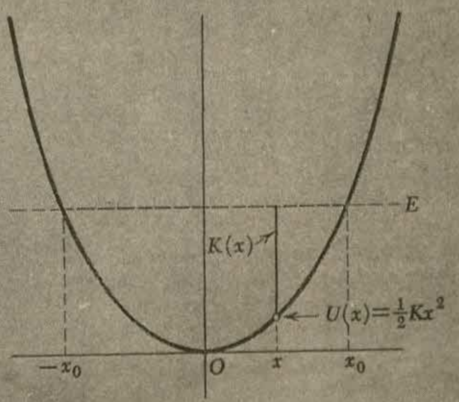


The total energy $E = K + U =$ const. Thus, given E, the motion can only occur between x' and x'', the "turning points." Between these points

$$K = \frac{Mv^2}{2} = E - U \geqq 0.$$



If E is increased, the turning points x' and x'' are, in general, changed. Now $K(x) = E - U(x)$ is greater. The motion can now also take place to the left of x''', if started at x'''.



The simple harmonic oscillator is in stable equilibrium at $x = 0$. At $x = \pm x_0$, $K = 0$.

---

† The derivatives involving the notation ∂ instead of d in (48) and (49) are called partial derivatives. When we have a function $f(x_1, x_2, \ldots)$ of two or more independent variables $x_1, x_2, \ldots$ and want to differentiate f with respect to one of them while holding the others constant, the process is called *partial differentiation*. Thus

$$\frac{\partial U}{\partial x} \equiv \left(\frac{\partial U}{\partial x}\right)_{y,z} = \lim_{\Delta x \to 0} \frac{U(x + \Delta x, y, z) - U(x, y, z)}{\Delta x}.$$

The subscripts y,z indicate explicitly that y and z are held constant.

where the sum is over all charges. The charge q_i is at the point $\mathbf{r}_i$. Suppose we bring up from infinity a test charge q. The force on q at $\mathbf{r}$ is

$$\mathbf{F}(\mathbf{r}) = q\mathbf{E}(\mathbf{r}). \tag{51}$$

The potential energy of the charge q in the field of all the other charges is given by (44*a*) as

$$U(\mathbf{r}) = \int_{\mathbf{r}}^{\infty} \mathbf{F}(\mathbf{r}') \cdot d\mathbf{r}' = q\int_{\mathbf{r}}^{\infty} \mathbf{E}(\mathbf{r}') \cdot d\mathbf{r}'. \tag{52}$$

We put primes on the integration variable $\mathbf{r}'$ to avoid confusion with the point $\mathbf{r}$ at which the potential is desired.

The *electrostatic potential* $\varphi(\mathbf{r})$ at $\mathbf{r}$ is defined as the *potential energy per unit positive charge,* in the field of force of all the other charges:

$$\varphi(\mathbf{r}) = \frac{U(\mathbf{r})}{q} = \int_{\mathbf{r}}^{\infty} \mathbf{E}(\mathbf{r}) \cdot d\mathbf{r}. \tag{53}$$

This is a very useful quantity. Notice that it is a scalar. It is most important to distinguish φ from the potential energy U. Beware also of the use in experimental work of the symbol V for both quantities, electrostatic potential and potential energy.

If we know $\mathbf{E}(\mathbf{r})$ everywhere, then we can find the electrostatic potential $\varphi(\mathbf{r})$ everywhere. (This assumes we decide on an origin for $\varphi(\mathbf{r})$.) It is convenient to work with $\varphi(\mathbf{r})$ because it is a scalar, whereas $\mathbf{E}(\mathbf{r})$ is a vector.

The voltage drop or *potential difference,* P.D., between two points $\mathbf{r}_1$ and $\mathbf{r}_2$ is defined as

$$\text{P.D.} = \varphi(\mathbf{r}_2) - \varphi(\mathbf{r}_1). \tag{54}$$

This is the change in the electrostatic potential energy of a unit positive charge when taken from $\mathbf{r}_1$ to $\mathbf{r}_2$. Thus for a charge q taken between these points the potential energy difference is

$$U(\mathbf{r}_2) - U(\mathbf{r}_1) = q[\varphi(\mathbf{r}_2) - \varphi(\mathbf{r}_1)]. \tag{54a}$$

The unit of electrostatic potential, or of potential difference, in the Gaussian CGS system is the *statvolt.* We saw in Chap. 4 that the unit of electric field intensity is called the statvolt/cm; but φ differs from $\mathbf{E}$ in dimensions by a length, and thus φ is measured in statvolts. It is also true that φ has the dimensions of [charge]/[length], so that the statcoul/cm is also a possible name for the unit of potential.

The *practical* unit of electrostatic potential or potential difference is the *volt.* The volt is used in everyday life and widely in the laboratory. The volt is *defined* so that

$$\left(\frac{c}{10^8}\right) \times \begin{pmatrix}\text{potential difference}\\ \text{in statvolts}\end{pmatrix} = \begin{pmatrix}\text{potential difference}\\ \text{in volts}\end{pmatrix}, \tag{54b}$$

where c is the speed of light in cm/sec. Or, approximately,

$$\boxed{(300) \times \begin{pmatrix}\text{potential difference}\\ \text{in statvolts}\end{pmatrix} \cong \begin{pmatrix}\text{potential difference}\\ \text{in volts}\end{pmatrix}.} \tag{54c}$$

Examples. Electrostatic field and potential; potential difference; statvolts; volts. (*a*) What is the magnitude of the electric field at a distance of 1 Å ($=10^{-8}$ cm) from a proton?

From Coulomb's law

$$E = \frac{e}{r^2} \cong \frac{5 \times 10^{-10}\ \text{statcoul}}{(1 \times 10^{-8}\ \text{cm})^2} \cong 5 \times 10^6\ \text{statvolt/cm}$$
$$\cong (300)(5 \times 10^6)\ \text{volt/cm} \cong 1.5 \times 10^9\ \text{volt/cm}.$$

The field is directed radially outward from the proton.

(*b*) What is the electrostatic potential at this distance?

By (53) we have, referred to $U = 0$ at $r = \infty$,

$$\varphi(r) = \int_r^\infty \frac{e}{r^2}\,dr = \frac{e}{r} \cong \frac{5 \times 10^{-10}\ \text{statcoul}}{1 \times 10^{-8}\ \text{cm}}$$
$$\cong 5 \times 10^{-2}\ \text{statvolts} \cong 15\ \text{volts}.$$

(*c*) What is the potential difference in volts between positions 1 Å and 0.2 Å from a proton?

The potential at 1×10^{-8} cm is 15 volts; at 0.2×10^{-8} cm it is 75 volts. The difference $75 - 15 = 60$ volts, or $\frac{60}{300} \cong 0.2$ statvolts.

(*d*) A proton is released from rest at a distance of 1 Å from another proton. What is the kinetic energy when the protons have moved infinitely far apart?

By conservation of energy we know that the kinetic energy must equal the original potential energy, which is

$$\frac{e^2}{r} \cong \frac{(4.8 \times 10^{-10}\ \text{statcoul})^2}{1 \times 10^{-8}\ \text{cm}} \cong 23 \times 10^{-12}\ \text{erg}.$$

If one proton is kept at rest while the other moves, the terminal velocity of the moving proton is given by (using conservation of energy)

$$\tfrac{1}{2}Mv^2 \cong 23 \times 10^{-12} \text{ erg};$$

$$v^2 \cong \frac{2 \times 23 \times 10^{-12} \text{ erg}}{1.67 \times 10^{-24} \text{ gm}} \cong 27 \times 10^{12} \text{ (cm/sec)}^2,$$

or

$$v \cong 5 \times 10^6 \text{ cm/sec.}$$

(*e*) A proton is accelerated from rest by a uniform electric field. The proton moves through a potential drop of 100 volts. What is its final kinetic energy? (Note that 100 volts $\cong$ 0.33 statvolts.)

The kinetic energy will be equal to the change in potential energy, which is $e\,\Delta\varphi$, or

$$(4.8 \times 10^{-10} \text{ statcoul})(0.33 \text{ statvolt}) \cong 1.6 \times 10^{-10} \text{ erg.}$$

Example. Electron volts. A convenient unit of energy in atomic physics and in nuclear physics is the electron volt (ev). One electron volt is defined as the potential energy difference of a charge *e* between two points having a potential difference of one volt. Thus

$$\begin{aligned} 1 \text{ electron volt} &\cong (4.80 \times 10^{-10} \text{ statcoul})(\tfrac{1}{300} \text{ statvolt}) \\ &\cong 1.60 \times 10^{-12} \text{ erg.} \end{aligned} \tag{54d}$$

An alpha particle (He^4 nucleus or doubly ionized helium atom) accelerated from rest through a potential difference of 1000 volts has a kinetic energy equal to

$$(2e)(1000 \text{ volts}) = 2000 \text{ electron volts,}$$

where

$$2000 \text{ ev} = (2 \times 10^3)(1.60 \times 10^{-12}) = 3.2 \times 10^{-9} \text{ erg.}$$

We have seen that the difference $(K_B - K_A)$ in kinetic energy of a particle between two points has the property that

$$K_B - K_A = \int_A^B \mathbf{F} \cdot d\mathbf{r}, \tag{55}$$

where **F** is the force acting on the particle. But we know from (44) that

$$U_B - U_A = -\int_A^B \mathbf{F} \cdot d\mathbf{r}, \tag{55a}$$

so that on adding (55) and (55*a*) we have

$$(K_B + U_B) - (K_A + U_A) = 0. \tag{56}$$

Thus *the sum of the kinetic and potential energies is a constant, independent of time and independent of the particle.* Rewriting (56), we have for a one-particle system the energy function

$$\boxed{E = \tfrac{1}{2}Mv^2(A) + U(A) = \tfrac{1}{2}Mv^2(B) + U(B),} \tag{57}$$

where E is a constant called the energy, or total energy, of the system.

Let us write down the generalization of (57) to a two-particle system in the field of an external potential:

$$\begin{aligned} E &\doteq K + U \\ &= \tfrac{1}{2}M_1v_1^2 + \tfrac{1}{2}M_2v_2^2 + U_1(\mathbf{r}_1) + U_2(\mathbf{r}_2) + U(\mathbf{r}_1 - \mathbf{r}_2) \\ &= \text{const.} \end{aligned} \tag{58}$$

The first term is the kinetic energy of particle 1; the second term is the kinetic energy of particle 2; the third and fourth terms are the potential energies of particles 1 and 2 due to an external potential; the fifth term is the potential energy due to the interaction between particles 1 and 2. Notice that $U(\mathbf{r}_1 - \mathbf{r}_2)$ is put in only once: if two particles interact, the interaction energy is mutual!

If particles 1 and 2 are protons in the earth's gravitational field, the energy E in (58) is

$$E = \frac{1}{2}M(v_1^2 + v_2^2) + Mg(x_1 + x_2) - \frac{GM^2}{r_{12}} + \frac{e^2}{r_{12}}, \tag{59}$$

where x is measured upward and $r_{12} = |\mathbf{r}_2 - \mathbf{r}_1|$. The last term is the Coulomb energy of the two protons; the next-to-last term is their gravitational energy. The Coulomb energy is repulsive; the gravitational energy is attractive. The ratio of the last two terms is

$$\frac{GM^2}{e^2} \approx \frac{(10^{-7})(10^{-48})}{10^{-19}} \approx 10^{-36},$$

showing that gravitational forces between protons are extremely weak in comparison with electrostatic forces.

Example. Linear restoring force. Calculate the potential energy $U(x)$ for a particle acted on by the linear restoring force

$F_{\text{int}} = -Cx$ due to a spring. (We do this as a one-dimensional example.)

From the relation (46) we have $F_{\text{int}} = -dU/dx$, so that on integration

$$U(x_2) - U(x_1) = -\int_{x_1}^{x_2} F_{\text{int}}\,dx = C\int_{x_1}^{x_2} x\,dx$$
$$= \frac{1}{2}C(x_2^2 - x_1^2), \qquad (60)$$

as in (28). In the present problem it is convenient to define the zero of potential energy at the origin, so that

$$U(x) = \tfrac{1}{2}Cx^2. \qquad (61)$$

Conservation (57) of the sum E of the potential and kinetic energies tells us that at any point x

$$E = \tfrac{1}{2}Mv^2 + \tfrac{1}{2}Cx^2 = \text{const.} \qquad (62)$$

When $x = 0$, the entire energy of the oscillator is kinetic; at the ends or turning points of the motion the velocity v is zero and the entire energy is potential. Therefore, if the maximum displacement is denoted by A, the total energy is

$$E = \tfrac{1}{2}CA^2. \qquad (63)$$

The maximum velocity is given by

$$v^2(\text{max}) = \frac{2E}{M} = \frac{CA^2}{M}. \qquad (64)$$

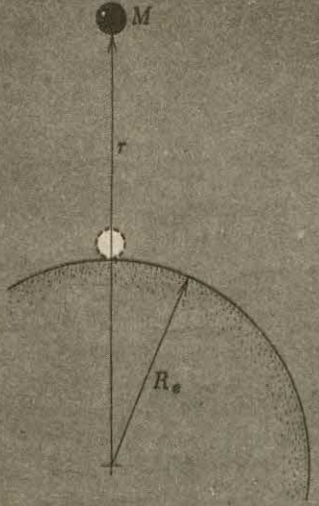
Example. Escape velocity from earth and from solar system. Calculate the initial velocity needed for a particle of mass M to escape (a) from the earth and (b) from the solar system. (Neglect the rotation of the earth.)

A particle of velocity v has a total energy, kinetic plus potential, of

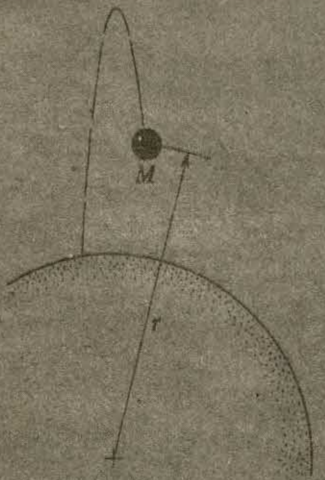
$$E = \frac{1}{2}Mv^2 - \frac{GM_eM}{R_e}, \qquad (65)$$

where G is the gravitational constant; M_e is the mass of the earth; and $R_e \cong 6.4 \times 10^8$ cm is the radius of the earth. The value of G is 6.670×10^{-8} dyne-cm²/gm²; the value of M_e is 5.98×10^{27} gm.

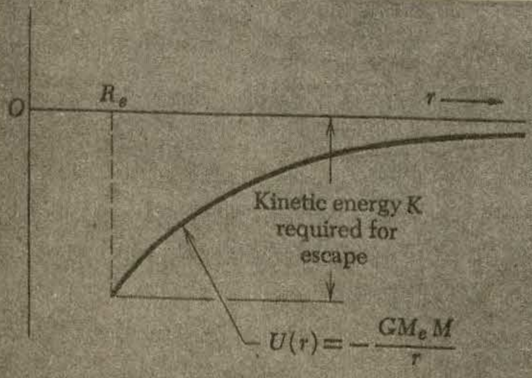
To reach an infinite distance from the earth with the least possible, that is zero, velocity, the total energy must be zero, because the kinetic energy is zero and the gravitational poten-



Consider the escape velocity required for a mass M to leave the earth's gravitational field starting from the surface.



Trajectory if the kinetic energy is too low for escape.



To reach infinity, $K \geqq -U = |U|$.

tial energy is also zero. This follows because $U(r) \to 0$ as $r \to \infty$. Thus in (65) E must be zero if the total energy of the particle is constant between launching and escape, whence the escape velocity v_e is given by

$$\frac{1}{2}Mv_e^2 = \frac{GM_eM}{R_e}; \qquad v_e = \sqrt{\frac{2GM_e}{R_e}}. \tag{66}$$

The acceleration g of gravity at the surface of the earth is GM_e/R_e^2, so that

$$v_e = \sqrt{2gR_e} \cong (2 \times 10^3 \times 6 \times 10^8)^{1/2} \approx 10^6 \text{ cm/sec}. \tag{67}$$

To escape from the pull of the sun alone a particle launched from the earth (at a distance R_{es} from the sun) will need an escape velocity

$$v_s = \sqrt{\frac{2GM_s}{R_{es}}} \approx \left[\frac{2 \times (7 \times 10^{-8}) \times (2 \times 10^{33})}{1.5 \times 10^{13}}\right]^{1/2}$$
$$\approx 4 \times 10^6 \text{ cm/sec}, \tag{68}$$

using the ratio $M_s/M_e = 3.3 \times 10^5$, and the value $R_{es} = 1.5 \times 10^{13}$ cm. For bodies launched from earth, escape from the solar system is more difficult than escape from the earth.

Example. Gravitational potential near surface of the earth. The gravitational potential energy of a body of mass M at a distance r from the center of the earth is, for $r > R_e$,

$$U(r) = -\frac{GMM_e}{r}, \tag{69}$$

where M_e is the mass of the earth. If R_e is the radius of the earth and x is the height above the earth's surface, show that

$$U \cong -MgR_e + Mgx, \tag{70}$$

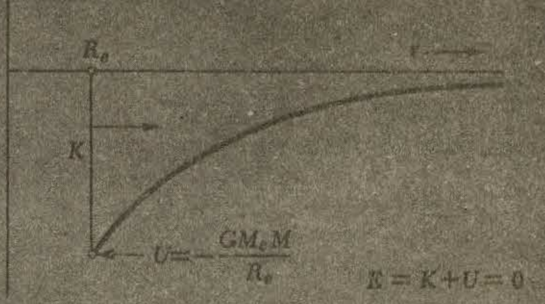
for $x/R_e \ll 1$. Here $g = GM_e/R_e^2 \approx 980$ cm/sec². From (46) and (70) we have for the gravitational force $F_G = -dU/dx = -Mg$, near the surface of the earth. This is the familiar result.

We have

$$U = -GMM_e\frac{1}{(R_e + x)}, \tag{71}$$

from (69) with $r = R_e + x$. Divide numerator and denominator by R_e:

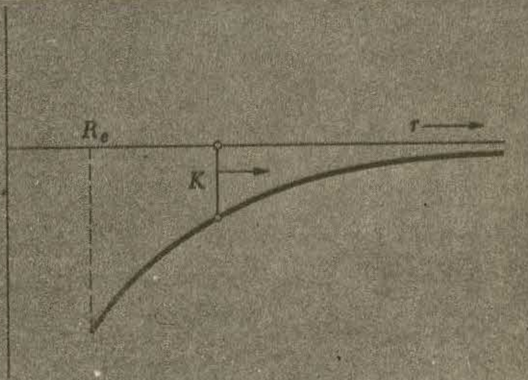
$$U = -\frac{GMM_e}{R_e} \times \frac{1}{(1 + x/R_e)}. \tag{72}$$



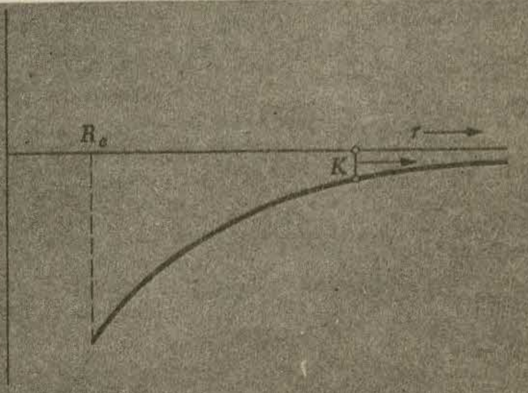
Here we see M launched from the earth's surface (radius R_e) with the minimum necessary kinetic energy

$$K = \frac{1}{2}Mv_e^2 = \frac{GM_eM}{R_e}.$$

The escape velocity from the earth is denoted by v_e.



Some time later U has increased and K has decreased as M goes further from the center of the earth.



Still later, K and $|U|$ have decreased further. Of course $K = -U$, as before.

Now for $x^2 < 1$ we have the important series expansion

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \cdots, \tag{73}$$

as in Dwight 9.04. This expansion is a special case for $n = -1$ of the binomial theorem:

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{2!}x^2 + \frac{n(n-1)(n-2)}{3!}x^3 + \cdots. \tag{74}$$

Thus (72) may be written as

$$U = -\frac{GMM_e}{R_e}\left(1 - \frac{x}{R_e} + \frac{x^2}{R_e{}^2} - \cdots\right). \tag{75}$$

Let $g = GM_e/R_e{}^2$; then

$$U = -MgR_e\left(1 - \frac{x}{R_e} + \frac{x^2}{R_e} - \cdots\right), \tag{76}$$

which reduces to (70) for $x \ll R_e$.

The approximation to (73) in which we keep the term in x of lowest order is

$$\frac{1}{1+x} \cong 1 - x, \quad (|x| \ll 1); \tag{77}$$

this is used very often in physics in circumstances where x^2 may be neglected in comparison with x. Another common and useful approximation is

$$(1+x)^n \cong 1 + nx, \quad (|nx| \ll 1), \tag{78}$$

which follows from (74).

Our statement

$$\text{Kinetic energy} + \text{potential energy} = \text{const} \tag{79}$$

of the law of conservation of energy will be generalized in Chap. 12 to include processes in which some or all of the mass is converted into energy. Such processes include most nuclear reactions. The necessary generalization is a natural consequence of the special theory of relativity. Equation (79) holds even when heat is produced in a process, because viewed on a microscopic scale, heat is just kinetic and potential energy of atoms, electrons, and molecules.

Further reading

PSSC, chaps. 21 (paragraph 8), 24 (paragraphs 1-5, and 25).

E. P. Wigner, "Symmetry and conservation laws," *Physics Today* **17**(3), 34 (1964).

Film list

"Energy and Work" (29 min) D. Montgomery (PSSC-MLA 0311). Discusses work done by constant force and by a variable force, and determination of the energy produced by such work.

"Elastic Collisions and Stored Energy" (28 min) J. Strickland (PSSC-MLA 0318). Quantitative demonstrations of the transformations between kinetic and potential energy in highly elastic collisions.

Problems

1. *Potential and kinetic energy—falling body.* (*a*) What is the potential energy of a mass of 1 kg at a height of 1 km above the earth? Express the answer in ergs, and refer the potential energy to the surface of the earth. *Ans.* 9.8×10^{10} ergs.

(*b*) What is the kinetic energy just as it touches the earth of a mass of 1 kg which is released from a height of 1 km? Neglect friction. *Ans.* 9.8×10^{10} ergs.

(*c*) What is the kinetic energy of the same mass when it has fallen halfway?

(*d*) What is the potential energy when it has fallen halfway? The sum of (*c*) and (*d*) should equal (*a*) or (*b*). Why?

2. *Potential energy above earth.* (*a*) What is the potential energy $U(R_e)$ of a mass of 1 kg on the surface of the earth, referred to zero potential energy at infinite distance? (Note that $U(R_e)$ is negative.) *Ans.* -6.23×10^{14} ergs.

(*b*) What is the potential energy of a mass of 1 kg at a distance of 10^5 km from the center of the earth, referred to zero potential energy at infinite distance? *Ans.* -3.98×10^{13} ergs.

(*c*) What is the work needed to move the mass from the surface of the earth to a point 10^5 km from the center of the earth?

3. *Electrostatic potential energy.* (*a*) What is the electrostatic potential energy of an electron and a proton at a separation of $1\ \text{Å} \equiv 10^{-8}$ cm, referred to zero potential energy at infinite separation? If charge is expressed in esu, the result will be in ergs. *Ans.* -2.3×10^{-11} erg.

(*b*) What is the electrostatic potential energy of two protons at the same separation? Pay special attention to the sign of the answer.

4. *Satellite in circular orbit.* (*a*) What is the centrifugal force on a satellite moving in a circular orbit about the earth at a distance r from the center of the earth? The velocity of the satellite relative to the center of the earth is v, and the mass is M.

(*b*) What is the gravitational force on the satellite?

(*c*) Express v in terms of r by equating the gravitational force to the centrifugal force.

5. *Moon—kinetic energy.* What is the kinetic energy of the moon relative to the earth? The relevant data are given in the table of constants inside the cover of this volume.

6. *Anharmonic spring.* A peculiar spring has the force law $F = -Dx^3$.

(*a*) What is the potential energy at x, referred to $U = 0$ at $x = 0$?

Ans. $\frac{1}{4}Dx^4$.

(*b*) How much work is done on the spring in stretching it slowly from 0 to x?

7. *Nonconservative force field?* Given the force field

$$\mathbf{F} = (y^2 - x^2)\hat{\mathbf{x}} + 3\,xy\hat{\mathbf{y}},$$

compute the line integral from the point (0,0) to the point (x_0,y_0) along the path made up of two straight sections (0,0) to $(x_0,0)$ and $(x_0,0)$ to (x_0,y_0). Compare with the result you get by taking the other two sides of the rectangle as the path of integration. Is the force conservative?

8. *Closest approach of protons.* Two protons each of energy 500 Mev approach each other from opposite directions. Here 1 Mev $\equiv 1 \times 10^6$ ev. If the only interaction is the electrostatic interaction e^2/r, how close could the protons come to each other? (Actually, experiments on scattering have shown that the range of nuclear interactions is of the order of 10^{-13} cm, and the observed radius of the proton is also about 10^{-13} cm, so that the behavior of two protons at these energies is not described very well by the electrostatic interaction.) *Ans.* 1.4×10^{-16} cm.

9. *Electron in bound orbit about proton.* Suppose that an electron moves in a circular orbit about a proton at a distance of 2×10^{-8} cm. Consider the proton to be at rest.

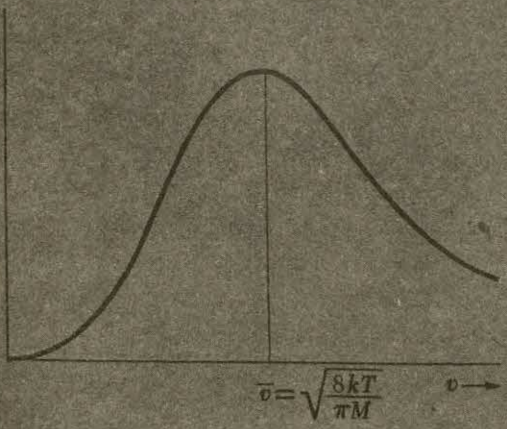
(*a*) Solve for the velocity of the electron by equating the centrifugal and electrostatic forces.

(*b*) What is the kinetic energy? potential energy? Give values both in ergs and in electron volts.

Ans. $K = 5.7 \times 10^{-12}$ erg $= 3.6$ ev; $U = -11.4 \times 10^{-12}$ erg $= -7.2$ ev.

(*c*) How much energy is needed to ionize the system—that is, to remove the electron to infinite distance with no final kinetic energy? Pay careful attention to the various signs.

10. *Escape of molecules from the atmosphere.* Compare the escape velocity from the earth with the root-mean-square thermal velocity of molecular oxygen. This r.m.s. thermal velocity is such that the kinetic energy is $(3/2)kT$; an appreciable fraction of molecules have much faster velocities. Here k is the Boltzmann constant and T is the absolute temperature, taken as 300°K. *Ans.* $v_{\text{rms}}/v_{\text{esc}} \cong 4 \times 10^{-2}$.



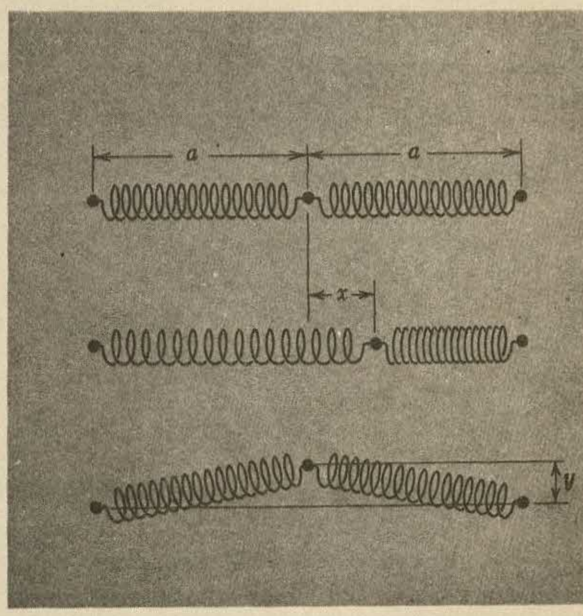
The velocity distribution function of molecules in a gas at temperature T has the form $v^2e^{-Mv^2/2kT}$, where k is the Boltzmann constant.

11. *Escape velocity from the moon.* Using $R_M = 1.7 \times 10^8$ cm and $M_M = 7.3 \times 10^{25}$ gm, find

(*a*) The gravitational acceleration at the surface of the moon.

(*b*) The escape velocity from the moon.

12. *Potential energy of pair of springs.* Two springs each of natural length a and stiffness constant C are fixed at points $(-a,0)$ and $(+a,0)$ and connected together at the other ends. In the following assume that either may expand or contract in length without buckling.



(*a*) Show that the potential energy of the system, for a displacement to (x,y) of the joined ends, is

$$U = \frac{C}{2}\{[(x+a)^2 + y^2]^{1/2} - a\}^2 + \frac{C}{2}\{[(a-x)^2 + y^2]^{1/2} - a\}^2.$$

(*b*) The potential energy depends on both x and y, and we must therefore use partial differentiation to evaluate the relevant forces. Remember that the partial derivative of a function $f(x,y)$ is taken by the usual rules of differentiation according to

$$\frac{\partial f(x,y)}{\partial x} = \frac{d}{dx} f(x;\ y = \text{const}); \qquad \frac{\partial f(x,y)}{\partial y} = \frac{d}{dy} f(x = \text{const};\ y).$$

Find the force component F_x and show that $F_x = 0$ for $r = 0$.

(*c*) Find F_y for $x = 0$. Check the signs carefully to make sure the answer makes sense.

(*d*) Sketch a graph of potential energy as a function of r in the xy plane, and find the equilibrium position.

13. *Nucleon-nucleon interaction.* The interaction between nucleons (protons and neutrons) may be represented to a fair degree of accuracy by the Yukawa potential,

$$U(r) = -\left(\frac{r_0}{r}\right) U_0 e^{-r/r_0},$$

where $U_0 \approx 50$ Mev and $r_0 \approx 1.5 \times 10^{-13}$ cm.

(*a*) Find the expression for the corresponding force $\mathbf{F}(r)$.

(*b*) At what approximate separation is the force 1 percent of its value at $r = r_0$? *Ans.* 6×10^{-13} cm.

14. *Time-of-flight mass spectrometer.* The operation of a time-of-flight mass spectrometer is based on the fact that the angular frequency of helical motion in a uniform magnetic field is independent of the initial velocity of the ion. In practice, the device produces a short pulse of ions and measures electronically the time of flight for one or more revolutions of the ions in the pulse.

(*a*) Show that the time of flight for N revolutions is approximately, for ions of charge e,

$$t \cong 650 \frac{NM}{B},$$

where t is in *microseconds, M* in *atomic mass units,* and B in *gauss.*

(*b*) Show that the gyroradius is approximately

$$R \cong \frac{145\sqrt{VM}}{B}\ \text{cm},$$

where V is the ion energy in electron volts.

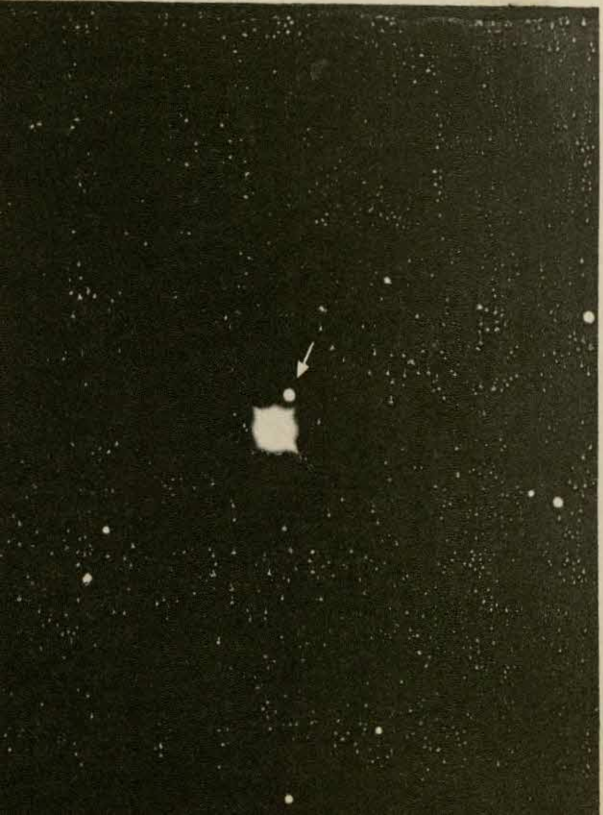
(*c*) Given an ion energy of 100 ev and a magnetic field of 1000 gauss, calculate the time of flight for 6 revolutions of singly ionized potassium K^{39}. *Ans.* 152 μsec.

15. *Electron beam in oscilloscope.* Electrons in an oscilloscope tube are accelerated from rest through a potential difference φ_a and pass between two electrostatic deflection plates. The plates, which have a length l and a separation d, sustain a potential difference φ_b with respect to each other. The screen of the tube is located at a distance L from the center of the plates. Use the relation $e\,\Delta\varphi = \frac{1}{2}mv^2$ between the accelerating potential and the velocity v.

(*a*) Derive an expression for the linear deflection D of the spot on the screen.

(*b*) Assume that $\varphi_a = 400$ volts; $\varphi_b = 10$ volts; $l = 2$ cm; $d = 0.5$ cm; $L = 15$ cm; what is this deflection? (Remember to change volts to statvolts.)

Neptune as seen in the Lick Observatory 120-in. reflector. The arrow points to Triton, a satellite of Neptune. (*Lick Observatory photograph*)



Historical Notes. Discovery of Ceres and of Neptune

(This discussion illustrates the accuracy of predictions based on classical mechanics.)

1. The first minor planet to be discovered was Ceres, found visually by Piazzi in Palermo in Sicily on the first day of the nineteenth century, January 1, 1801. Piazzi observed its motions for a few weeks, then became ill and lost track of it. A number of scientists calculated its orbit from the limited number of positions observed by Piazzi, but only the orbit computed by Gauss was accurate enough to predict where it might be the next year. On January 1, 1802, the planet Ceres was rediscovered by Olbers at an angular distance of only 30′ from the predicted location. As more observations accumulated, Gauss and others were able to improve the characteristics of the calculated orbit, and by 1830 the position of the planet was only 8″ from its predicted location. By including the major perturbations of the orbit of Ceres due to Jupiter, Enke found that he could reduce the residual error to an average of 6″ per year. Later computations, taking perturbations into account more accurately, produced predictions which disagreed with observations only by some 30″ after 30 years.

Accounts of the discovery are given in vol. 12 of the *Philosophical Magazine,* (1802); see the papers by Piazzi (p. 54), Von Zach (p. 62), Tilloch (p. 80), and Lalande (p. 112). It is amusing to find that a society of eminent astronomers of Europe was organized at Lilienthal on September 21, 1800 for "the express purpose of searching out this planet supposed to exist between Mars and Jupiter. . . . The plan of the society was to divide the whole zodiac among the twenty-four members. . . ." Thanks to the postal delays of the Napoleonic war, the invitation to participate in team research was not transmitted to Piazzi until after he had made the discovery. Other accounts of the discovery of Ceres are found in the *Astronomisches Jahrbuch* 1804/5. Calculations by Gauss are in vol. 6 of his *Werke*, pp. 199–211.

The astronomical tradition started in Palermo by the Abbot Piazzi is believed to have reached Lampedusa (hero of the novel *The Leopard*, written by his descendant) by way of the Abbot Pirrone, who was Lampedusa's spiritual advisor and astronomical assistant.

2. During the first half of the nineteenth century, as the precision of observations and of theory improved, it was found that the planet Uranus was not moving according to the law of gravitation and the conservation of energy and angular momentum. The planet erratically accelerated and decelerated by small but very significant amounts. There was no way of explaining this behavior on the basis of the known properties of the solar system and the laws of physics. Finally, in 1846, Leverrier and Adams independently discovered that the postulation of a hypothetical new planet of a certain mass, and of a certain orbit exterior to that of Uranus, would completely explain the observed anomalous motion.† They solved their equations for the location of this unknown planet and after only a half hour of search, the new planet, named Neptune, was found by Galle only 1° from the predicted location.‡ Present predictions of the positions of the major planets agree within a few seconds of arc with observations, even after an extrapolation of many years. The accuracy seems completely dependent upon the completeness of treatment of the various perturbing effects.

---

† "I proved it was not possible to account for the observations of this planet [Uranus] by the theory of universal gravitation if the planet were subject only to the combined action of the sun and of the known planets. But all the observed anomalies can be explained to the smallest detail by the influence of a new [undiscovered] planet beyond Uranus. . . . We predict [August 31, 1846] the following position for the new planet on 1 January 1847: True heliocentric longitude 326° 32′." U. J. Le Verrier, *Compt. Rend.* **23**, 428 (1846).

‡ "I wrote to M. Galle on the 18th of September to ask his cooperation; this able astronomer saw the planet the same day [September 23, 1846] he received my letter. . . . [Observed] heliocentric longitude 327°24′ reduced to 1 January 1847. . . . Difference [observed and theory] 0°52′." Le Verrier, *loc cit.*, p. 657.

"M. Le Verrier saw the new planet without having need to glance a single time at the sky; he saw it at the *end of his pen;* he determined by the sole power of calculation the position and size of a body situated well beyond the then known limits of our planetary system. . . ." Arago, *loc. cit.*, p. 659.

For an introduction to a magnificent controversy about the discovery, see pp. 741–754 of the same volume of the *Compt. Rend.* (Paris); see also M. Grosser, *The discovery of Neptune* (Harvard, Cambridge, Mass., 1962).

Chapter 6

Conservation of linear momentum 165
Center of mass 167
Examples: Collision of particles which stick together 168
Transverse momentum components 170
Collision of particles with internal excitations 170
Deflection of heavy particle by light particle 170
Satellite problem 173
Space-vehicle problem 173
Conservation of angular momentum 175
Examples: Internal torques 177
Proton scattering by heavy nucleus 179
Rotational invariance 180
Example: Angular acceleration accompanying contraction 181
Shape of the Galaxy 182
Angular momentum of the solar system 184
Intrinsic angular momentum of elementary particles 185
Further reading 185
Film list 185
Problems 185
Advanced Topic: Meteoroid's encounter with the atmosphere 188

Conservation of Linear and Angular Momentum

Conservation of Linear Momentum

In Chap. 3 we considered systems for which Galilean invariance was valid, and we showed that the conservation of linear momentum of a system of interacting particles is a necessary consequence of Galilean invariance and conservation of energy, provided that no external force acts. The conservation of linear momentum, a law accurately verified by experiment, is an essential part of the "classical package" as discussed previously. We want in this chapter to define the center of mass, and to consider collision processes as viewed from a reference frame in which the center of mass is at rest.

We shall first develop the form taken by the potential energy of a system when Galilean invariance holds. Consider two particles in one dimension. The coordinates are x_1 and x_2. The potential energy

$$U(x_1,x_2)$$

depends only on the positions of the two particles. To have Galilean invariance U must be invariant (unchanged) under a uniform translation b of all particles:

$$U(x_1,x_2) = U(x_1 + b;\ x_2 + b), \tag{1}$$

for a translation of every particle by b. What does this relation mean? Physically, it means that the potential energy of the system is the same at the displaced position as at the initial position, so that no work is done in the displacement. Mathematically, it means that b does not appear in the development of the expression on the right-hand side of (1).

For example, if

$$U(x_1,x_2) = (x_1 - x_2)^2, \tag{2}$$

it follows that

$$\begin{aligned} U(x_1 + b; x_2 + b) &= (x_1 + b - x_2 - b)^2 \\ &= (x_1 - x_2)^2 = U(x_1,x_2), \end{aligned} \tag{3}$$

independent of b. Thus the form (2) for $U(x_1,x_2)$ is invariant under any translation b. In general, if $U(\mathbf{r}_1,\mathbf{r}_2)$ is a function of only $\mathbf{r}_1 - \mathbf{r}_2$, it will be translationally invariant. Notice that, for example, $U = x_1$ is not invariant, for under translation by b it becomes $x_1 + b$. Thus momentum is not conserved if $U = x_1$.

We know in one dimension that the force is the negative of dU/dx; thus the forces on particles 1 and 2 are

$$F_1 = -\frac{\partial U}{\partial x_1}; \qquad F_2 = -\frac{\partial U}{\partial x_2}. \tag{4}$$

In taking $\partial U/\partial x_1$ we are to hold x_2 constant; in taking $\partial U/\partial x_2$ we hold x_1 constant. If U is a function only of

$$u = x_1 - x_2, \tag{5}$$

then by the rules of differentiation,

$$\frac{\partial U}{\partial x_1} = \frac{dU}{du}\frac{\partial u}{\partial x_1} = \frac{dU}{du}; \qquad \frac{\partial U}{\partial x_2} = \frac{dU}{du}\frac{\partial u}{\partial x_2} = -\frac{dU}{du}, \tag{6}$$

so that

$$\frac{\partial U}{\partial x_1} = -\frac{\partial U}{\partial x_2}; \qquad \text{or} \qquad F_1 = -F_2, \tag{7}$$

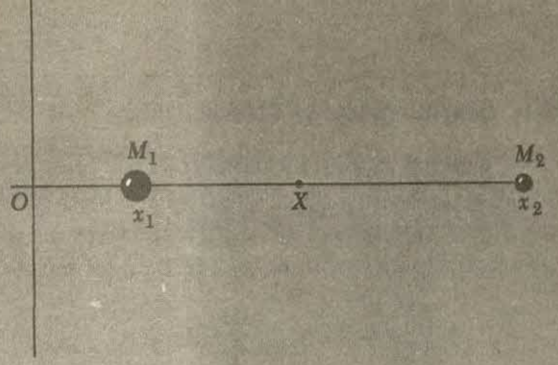
as expected. It follows that the total force $F_1 + F_2$ on two particles which interact only with each other is zero. If the total force is zero, then by Newton's law the total momentum is constant.

It is easy to generalize this argument to N particles. The N-particle potential energy $U(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ is translationally invariant if it can be written as a function of only the interparticle distances

$$\mathbf{r}_1 - \mathbf{r}_2, \quad \mathbf{r}_1 - \mathbf{r}_3, \ldots, \mathbf{r}_1 - \mathbf{r}_N, \quad \mathbf{r}_2 - \mathbf{r}_3, \ldots, \mathbf{r}_{N-1} - \mathbf{r}_N.$$

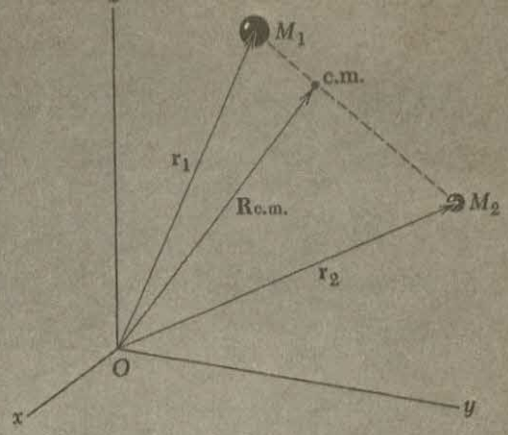
The Coulomb and the gravitational interactions may be written in this form.

Note that *force* may be invariant under translation without having momentum conserved. What is important is that the



Two masses M_1 and M_2 at positions x_1 and x_2 on the x axis have a center of mass located at

$$X = \frac{M_1x_1 + M_2x_2}{M_1 + M_2}.$$



For two masses M_1 and M_2 at arbitrary positions $\mathbf{r}_1$ and $\mathbf{r}_2$,

$$\mathbf{R}_{\text{c.m.}} = \frac{M_1\mathbf{r}_1 + M_2\mathbf{r}_2}{M_1 + M_2}.$$

potential energy be invariant. Momentum is constant only when the potential energy is translationally invariant.

Center of Mass. Relative to a fixed origin O, the position $\mathbf{R}_{\text{c.m.}}$ of the center of mass (c.m.) of a system of N particles is defined as

$$\mathbf{R}_{\text{c.m.}} = \frac{\sum_{n=1}^{N} \mathbf{r}_n M_n}{\sum_{n=1}^{N} M_n}. \tag{8}$$

For a two-particle system

$$\mathbf{R}_{\text{c.m.}} = \frac{\mathbf{r}_1 M_1 + \mathbf{r}_2 M_2}{M_1 + M_2}. \tag{9}$$

Differentiate (8) with respect to time:

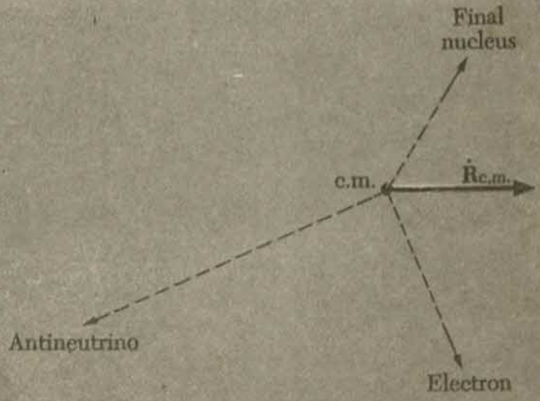
$$\dot{\mathbf{R}}_{\text{c.m.}} = \frac{\sum_n \dot{\mathbf{r}}_n M_n}{\sum_n M_n} = \frac{\sum_n \mathbf{v}_n M_n}{\sum_n M_n}; \tag{10}$$

but $\sum_n \mathbf{v}_n M_n$ is just the total momentum of the system. In the absence of external forces the total momentum is constant, so that

$$\dot{\mathbf{R}}_{\text{c.m.}} = \text{const.} \tag{11}$$

This is a remarkable property of the center of mass: the *velocity of the center of mass is constant in the absence of external forces.* This is true, for example, for a radioactive nucleus which decays in flight.

In the absence of external forces the velocity of the center of mass is constant. Here a radioactive nucleus with velocity $\dot{\mathbf{R}}_{\text{c.m.}}$ is about to decay.



The nucleus decays into three particles which go off in separate directions. However, the velocity of the center of mass of these three particles remains unchanged.

It is a simple matter to show from (10) that the acceleration of the center of mass is determined by the total external force acting on the system of particles. If $\mathbf{F}_n$ is the force on particle n, then on differentiating (10) with respect to time we have

$$\left(\sum_n M_n\right)\ddot{\mathbf{R}}_{\text{c.m.}} = \sum_n (M_n \dot{\mathbf{v}}_n) = \sum_n \mathbf{F}_n = \mathbf{F}_{\text{ext}}, \tag{12}$$

where on the right the interparticle internal forces drop out (for Newtonian forces) in the sum $\sum_n \mathbf{F}_n$ over all particles. We recall that for a Newtonian force the force on particle i due to particle j is equal but opposite to the force on j due to i.

We are now going to illustrate the usefulness of the center of mass by working out several important collision problems.

Example. Collision of particles which stick together. Consider the collision of two particles of mass M_1 and M_2 which stick together on collision. Let M_2 be at the origin and at rest before the collision, and let $\mathbf{r}_1 = v_1 t\,\hat{\mathbf{x}}$ describe the motion of M_1 before the collision.

(a) Describe the motion of $M = M_1 + M_2$ after the collision. Regardless of whether the collision is elastic or inelastic, the total momentum is unchanged in a collision. The collision considered here is inelastic. The initial x component of momentum is $M_1 v_1$; the final x component of momentum is $(M_1 + M_2)v$. The other components are zero. By conservation of momentum we have

$$M_1 v_1 = (M_1 + M_2)v, \tag{13}$$

so that the final velocity v is given by

$$v = \frac{M_1}{M_1 + M_2} v_1, \tag{14}$$

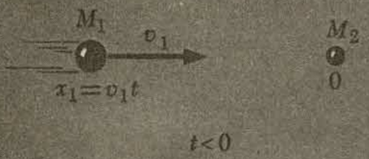
and, because the particles are stuck together,

$$\mathbf{R}_{\text{c.m.}} = vt\,\hat{\mathbf{x}} \qquad (t > 0) \tag{15}$$

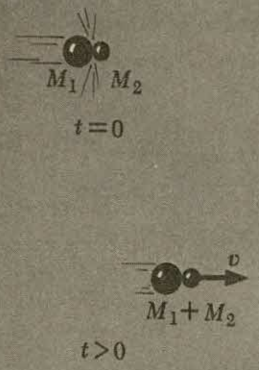
describes the motion of the system after the collision. According to (11) this same relation must describe the motion of the center of mass at all times, before or after the collision:

$$\mathbf{R}_{\text{c.m.}} = vt\,\hat{\mathbf{x}} = \frac{M_1}{M_1 + M_2} v_1 t\,\hat{\mathbf{x}}, \tag{16}$$

using (13).



Even in an inelastic collision, momentum must be conserved. Consider a collision in which the particles stick together. Before the collision $p_x = M_1 v_1$.

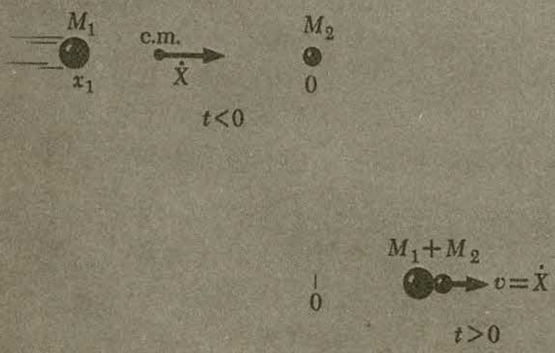


After the collision

$$p_x = M_1 v_1 = (M_1 + M_2)v,$$

so

$$v = \frac{M_1 v_1}{M_1 + M_2} < v_1.$$



$$X = \frac{M_1 x_1 + M_2 x_2}{M_1 + M_2}, \quad \text{so } \dot{X} = \frac{M_1 v_1}{M_1 + M_2}.$$

$\dot{X}$ is unchanged by the collision.

(*b*) What is the ratio of the kinetic energy after the collision to the initial kinetic energy? The kinetic energy K_f after the collision is

$$K_f = \frac{1}{2}(M_1 + M_2)\frac{M_1^2}{(M_1 + M_2)^2}v_1^2 = \frac{M_1^2 v_1^2}{2(M_1 + M_2)}. \tag{17}$$

The initial kinetic energy K_i is equal to $\frac{1}{2}M_1v_1^2$, so that

$$\frac{K_f}{K_i} = \frac{M_1}{M_1 + M_2}. \tag{17a}$$

The remainder of the energy appears in the internal excitations of the composite system after the collisions. When a meteorite M_1 strikes and sticks to the earth M_2, essentially all the kinetic energy of the meteorite appears as heat in the earth, because $M_1 \ll M_1 + M_2$.

(*c*) Describe the motion before and after the collision in the reference frame in which the center of mass is at rest. (Such a reference is called the center-of-mass system.)

The position of the center of mass of the system is given from (9) as

$$\mathbf{R}_{\text{c.m.}} = \frac{M_1 v_1 t\,\hat{\mathbf{x}}}{M_1 + M_2}. \tag{18a}$$

The velocity of the center of mass is given by

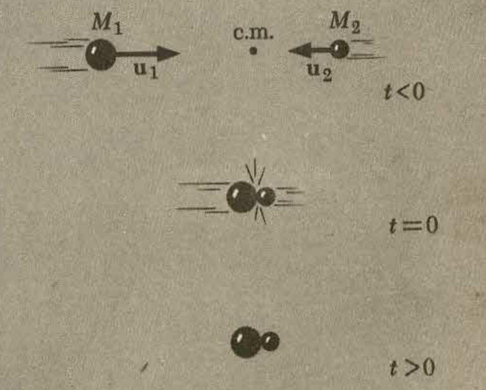
$$\mathbf{V} = \frac{d}{dt}\mathbf{R}_{\text{c.m.}} = \frac{M_1}{M_1 + M_2}v_1\,\hat{\mathbf{x}}. \tag{18b}$$

In the center-of-mass reference frame the initial velocity $\mathbf{u}_1$ of particle 1 is

$$\mathbf{u}_1 = v_1\,\hat{\mathbf{x}} - \mathbf{V} = \left(1 - \frac{M_1}{M_1 + M_2}\right)v_1\,\hat{\mathbf{x}} = \frac{M_2}{M_1 + M_2}v_1\,\hat{\mathbf{x}}. \tag{18c}$$

In the center-of-mass reference frame the initial velocity $\mathbf{u}_2$ of particle 2 is

$$\mathbf{u}_2 = -\mathbf{V} = -\frac{M_1}{M_1 + M_2}v_1\,\hat{\mathbf{x}}. \tag{18d}$$



In the center-of-mass reference frame the velocities of M_1 and M_2 before collision are $\mathbf{u}_1$, $\mathbf{u}_2$. After collision $(M_1 + M_2)$ is at rest.

Now, on colliding, the particles stick together; the new combined particle has mass $M_1 + M_2$ and must be at rest in the center-of-mass system. Relative to the laboratory system the new particle has the velocity $\mathbf{V}$ given by (18*b*), which is exactly the velocity (14) as found by the earlier argument.

Example. Transverse momentum components. Two particles of equal mass move initially on paths parallel to the x axis and collide. After the collision one of the particles is observed to have a particular value $v_y(1)$ of the y component of the velocity. What is the y component of the velocity of the other particle after the collision? Recall that each component x, y, or z of the total linear momentum is conserved separately.

Before the collision the particles were moving along the x axis, so that the total y component of the momentum is zero. By momentum conservation the total y component of momentum must also be zero after the collision, so that

$$M[v_y(1) + v_y(2)] = 0, \tag{19}$$

whence

$$v_y(2) = -v_y(1). \tag{20}$$

We cannot calculate $v_y(1)$ itself without specifying the initial trajectories and the details of the forces during the collision process.

Example. Collision of particles with internal excitations. Two particles of equal mass and equal but opposite velocities $\pm\mathbf{v}_i$ collide. What are the velocities after the collision?

The center of mass is at rest and must remain at rest, so that the final velocities $\pm\mathbf{v}_f$ are equal but opposite. If the collision is elastic, the conservation of energy demands that the final speed v_f equal the initial speed v_i. If one or both particles are excited internally by the collision, then $v_f < v_i$ by conservation of energy. If one or both particles initially are in excited states of internal motion and on collision they give up their excitation energy into kinetic energy, then v_f will be larger than v_i.

Example. Deflection of heavy particle by light particle. This is a famous problem. A particle of mass M_1 collides elastically with a particle of mass M_2 which initially is at rest in the laboratory frame of reference. The trajectory of M_1 is deflected through an angle θ_1 by the collision. The maximum possible value of the scattering angle θ_1 is determined by the laws of conservation of energy and momentum, independent of the details of the interaction between the particles. Our problem is to find $(\theta_1)_{\max}$. We shall see that it is convenient at one stage in the calculation to view the collision from the frame of reference in which the center of mass is at rest.

We denote the initial velocities in the laboratory frame by

$$\mathbf{v}_1 = v_1\hat{\mathbf{x}}; \qquad \mathbf{v}_2 = 0, \tag{21}$$

and the final velocities (after the collision) by $\mathbf{v}_1'$ and $\mathbf{v}_2'$. The law of conservation of energy requires that in an elastic collision the total kinetic energy before the collision equal the total kinetic energy after the collision. Thus

$$\tfrac{1}{2}M_1v_1^2 = \tfrac{1}{2}M_1v_1'^2 + \tfrac{1}{2}M_2v_2'^2, \tag{22}$$

noting the initial condition $v_2 = 0$. The law of conservation of momentum applied to the x component of momentum requires that

$$M_1v_1 = M_1v_1'\cos\theta_1 + M_2v_2'\cos\theta_2. \tag{23}$$

The entire collision event can be considered to lie in the xy plane, provided that only two particles are involved. The law of conservation of momentum applied to the y component of momentum requires that

$$0 = M_1v_1'\sin\theta_1 + M_2v_2'\sin\theta_2, \tag{24}$$

because initially the y component of momentum was zero.

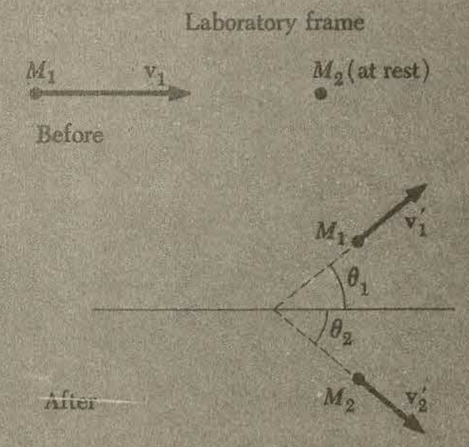
It is perfectly possible, but a little tedious, to solve (22) to (24) simultaneously for whatever quantities interest us. These equations express the entire content of the conservation laws. But it is considerably neater and more informative to view the collision in the center-of-mass reference frame. First we find the velocity $\mathbf{V}$ of the center of mass relative to the laboratory frame. The position of the center of mass is defined by

$$\mathbf{R}_{\text{c.m.}} = \frac{M_1\mathbf{r}_1 + M_2\mathbf{r}_2}{M_1 + M_2}. \tag{25}$$

The velocity $\mathbf{V}$ of the center of mass is given by

$$\dot{\mathbf{R}}_{\text{c.m.}} = \frac{M_1\dot{\mathbf{r}}_1 + M_2\dot{\mathbf{r}}_2}{M_1 + M_2}; \qquad \mathbf{V} = \frac{M_1}{M_1 + M_2}\mathbf{v}_1, \tag{26}$$

where we have expressed the result in terms of the velocities $\mathbf{v}_1$, $\mathbf{v}_2$ before the collision, with $\mathbf{v}_2 = 0$. Note that (26) is the same as (18b), but written now in three dimensions.



A collision between M_1 and M_2 need not be confined to one dimension. In the laboratory frame M_2 is at rest before the collision.

We denote by $\mathbf{u}_1$, $\mathbf{u}_2$ the initial velocities in the reference frame in which the center of mass is at rest; the final velocities in this reference frame are $\mathbf{u}_1'$, $\mathbf{u}_2'$. We have the following relations between the velocities in the laboratory and center-of-mass systems:

$$\mathbf{v}_1 = \mathbf{u}_1 + \mathbf{V}; \qquad \mathbf{v}_2 = \mathbf{u}_2 + \mathbf{V};$$
$$\mathbf{v}_1' = \mathbf{u}_1' + \mathbf{V}; \qquad \mathbf{v}_2' = \mathbf{u}_2' + \mathbf{V}. \tag{27}$$

We can make an educated guess at the kind of solution allowed by the conservation laws. Observe first that any collision for which the speeds

$$u_1 = u_1' \qquad \text{and} \qquad u_2 = u_2' \tag{28}$$

will conserve energy, because the individual particle kinetic energies are unchanged. Further, momentum will be conserved in the center-of-mass system if the scattering angle of particle 1 is equal to the scattering angle of particle 2, i.e., if the trajectories of the two particles are collinear. If these scattering angles (as viewed in the center-of-mass system) were not equal, then the center of mass could not be at rest after the collision. But we know that the center of mass remains at rest if no external forces act.

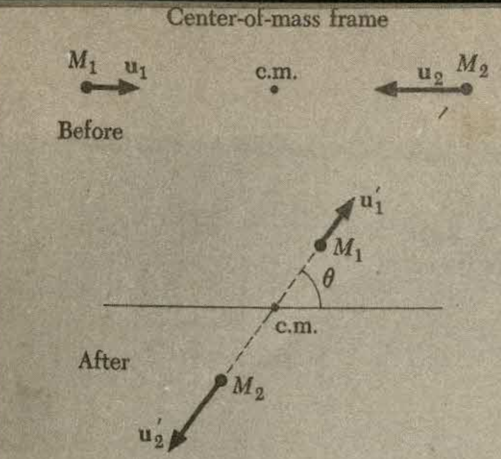
The kinematics of the collision in the center-of-mass system is seen to be trivially simple. All scattering angles $\theta_{\text{c.m.}}$ are allowed by the conservation laws. This is not true for θ_1, θ_2 in the laboratory system: consider a heavy particle incident on a light particle initially at rest. We know intuitively that the heavy particle will not bounce backward as a result of the collision. What are the restrictions?

Let us return to the laboratory system. We form, writing θ for $\theta_{\text{c.m.}}$ for convenience,

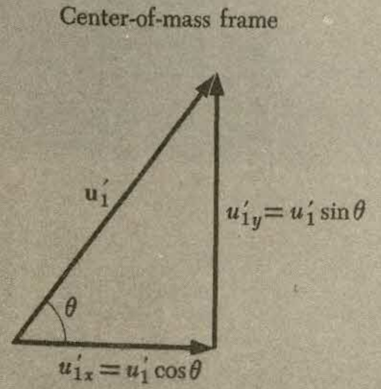
$$\tan\theta_1 = \frac{\sin\theta_1}{\cos\theta_1} = \frac{v_1'\sin\theta_1}{v_1'\cos\theta_1} = \frac{u_1'\sin\theta}{u_1'\cos\theta + V}, \tag{29}$$

where we have used the fact that the y component of the final velocity of particle 1 is identical in the two reference frames. Further, $u_1 = u_1'$ by the assumption (28), so that

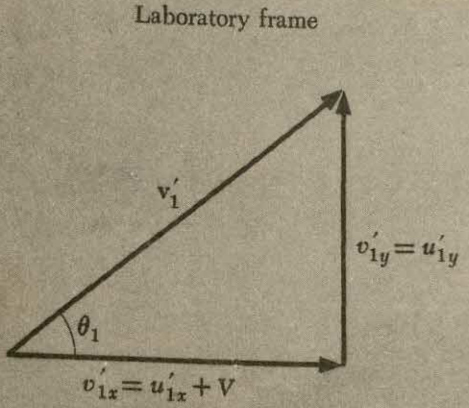
$$\tan\theta_1 = \frac{\sin\theta}{\cos\theta + V/u_1}. \tag{30}$$



In the center-of-mass frame M_1 and M_2 must go off in opposite directions after collision. All angles $0 \leqq \theta \leqq \pi$ are possible, and $|\mathbf{u}_1'| = |\mathbf{u}_1|$, $|\mathbf{u}_2'| = |\mathbf{u}_2|$.



The final velocity $\mathbf{u}_1'$ of M_1 in the center-of-mass frame is resolved into x and y components in the figure.



In the laboratory frame the x and y components of $\mathbf{v}_1'$ are as shown. Evidently

$$\tan\theta_1 = \frac{\sin\theta}{\cos\theta + V/u_1} = \frac{\sin\theta}{\cos\theta + M_1/M_2}.$$

Now (26) and (27) may be combined to give the relations

$$\mathbf{V} = \frac{M_1}{M_1 + M_2}(\mathbf{u}_1 + \mathbf{V}); \qquad \mathbf{V} = \frac{M_1}{M_2}\mathbf{u}_1, \tag{31}$$

whence (30) becomes

$$\tan\theta_1 = \frac{\sin\theta}{\cos\theta + M_1/M_2}. \tag{32}$$

We want to know the value of $(\theta_1)_{\text{max}}$. This can be found graphically from (32) or by using calculus to determine the maximum of tan θ_1 as a function of θ. We see by inspection that for $M_1 > M_2$ the denominator can never be zero, and $(\theta_1)_{\text{max}}$ must then be less than $\frac{1}{2}\pi$. If $M_1 = M_2$, then $(\theta_1)_{\text{max}} = \frac{1}{2}\pi$. If $M_1 < M_2$, any value of θ_1 is allowed.

Example. Satellite problem. A satellite in force-free space sweeps up stationary interplanetary debris at a rate $dM/dt = \alpha v$, where M is the mass and v the speed of the satellite; α is a constant. What is the deceleration?

From Newton's second law

$$\mathbf{F} = \frac{d}{dt}(M\mathbf{v}) = \dot{M}\mathbf{v} + M\dot{\mathbf{v}} = 0, \tag{33}$$

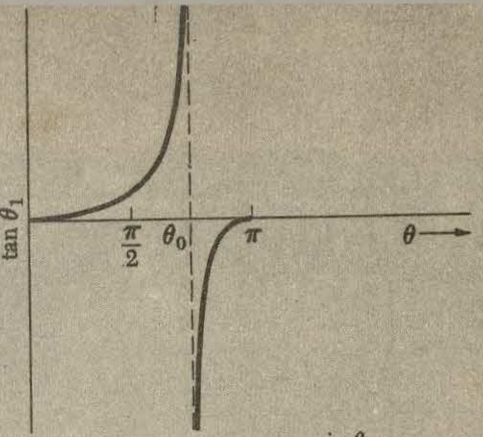
or, for motion in one dimension with $\dot{M} = \alpha v$,

$$\dot{v} = -\frac{\alpha v^2}{M}. \tag{34}$$

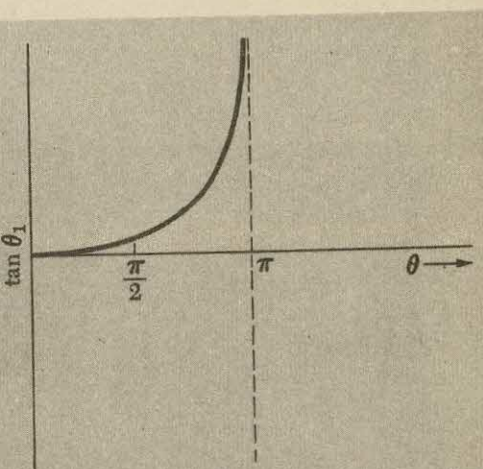
(This has been written for $v \geqq 0$.) Is there another way of looking at this problem?

Example. Space-vehicle problem. A space vehicle ejects fuel at a velocity $-\mathbf{V}_0$ relative to the vehicle; the rate of change of mass of the vehicle is $\dot{M} = -\alpha$, a constant. Set up and solve the equation of motion of the space vehicle, neglecting gravity.

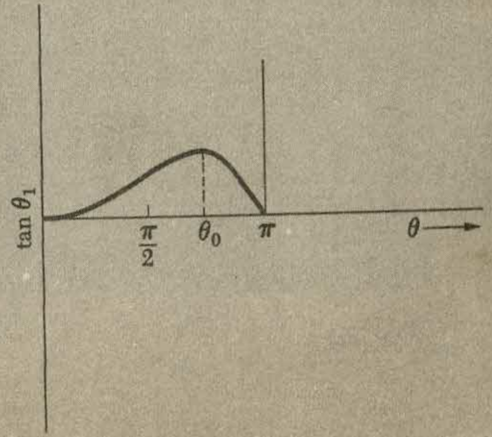
Let the velocity of the vehicle at time t be $\mathbf{v}$. The velocity of the fuel viewed in the laboratory (*not* in the vehicle) frame of reference (an inertial frame) is $-\mathbf{V}_0 + \mathbf{v}$. We assume that $\mathbf{V}_0$ and $\mathbf{v}$ are parallel, so that the problem reduces to a one-dimensional problem.



For $M_1 < M_2$, $\tan\theta_1 = \dfrac{\sin\theta}{\cos\theta + M_1/M_2}$ goes to infinity at $\theta = \theta_0 = \cos^{-1}(-M_1/M_2)$. All angles $0 \leqq \theta_1 \leqq \pi$ are possible.



For $M_1 = M_2$, tan θ_1 goes to infinity at $\theta = \pi$. Thus all angles $0 \leqq \theta_1 \leqq \pi/2$ are possible as roots of the equation for tan θ_1.



For $M_1 > M_2$, tan θ_1 does not go to infinity. Thus $0 \leqq \theta_1 \leqq \sin^{-1}(M_2/M_1) < \pi/2$.

The force on the satellite due to the ejection of fuel is equal but opposite to the momentum per unit time of the ejected fuel. Viewed in the laboratory frame of reference the momentum of the fuel ejected per unit time is $|\dot{M}|(-V_0 + v)$, or $\alpha(-V_0 + v)$, so that the force on the satellite is

$$F = -\alpha(-V_0 + v) = \alpha(V_0 - v), \tag{35}$$

still viewed in the laboratory frame of reference. This force is equal to $\dot{M}v + M\dot{v}$ for the satellite, so that the equation of motion is

$$\dot{M}v + M\dot{v} = \alpha(V_0 - v), \tag{36}$$

or

$$M\dot{v} = \alpha V_0, \tag{37}$$

where

$$M(t) = M_0 - \alpha t. \tag{38}$$

(This last equation assures that $\dot{M} = -\alpha$, as assumed.) Then (36) becomes

$$-\alpha v + (M_0 - \alpha t)\frac{dv}{dt} = \alpha(V_0 - v), \tag{39}$$

whence, on rewriting

$$dv = \frac{(\alpha V_0/M_0)}{1 - (\alpha t/M_0)}\,dt. \tag{39a}$$

On integrating between $t = 0$ and t, and v_0 and v, we find

$$v = -V_0 \log\left(1 - \frac{\alpha t}{M_0}\right) + v_0, \tag{40}$$

where v_0 is the initial velocity of the vehicle.

Thus the change in velocity in a time t is

$$v - v_0 = -V_0 \log\left(1 - \frac{t}{t_0}\right) = V_0 \log\frac{t_0}{t_0 - t}. \tag{41}$$

Here

$$t_0 \equiv \frac{M_0}{\alpha} \tag{42}$$

is the extrapolated time at which the entire mass of the vehicle would be expelled. Of course the vehicle is not all fuel; the engine cuts off at a time which is usually 85 to 90 percent of t_0. That is, 85 to 90 percent of the initial mass of the vehicle is fuel. The result (41) exhibits the advantage of a high-velocity (high V_0) propellant. The most efficient possible propellant would be photons, for which $V_0 = c$. Plot (41) in dimensionless form, using as variables $(v - v_0)/V_0$ and t/t_0. What kind of graph paper is best suited for this plot?

Conservation of Angular Momentum

The *angular momentum* **J** of a single particle referred to an arbitrary fixed point (fixed in an *inertial* reference frame) as origin is defined as

$$\boxed{\mathbf{J} \equiv \mathbf{r} \times \mathbf{p} \equiv \mathbf{r} \times M\mathbf{v},} \tag{43}$$

where **p** is the linear momentum. The units of angular momentum are gm-cm²/sec or erg-sec. The component of **J** along any line (or axis) passing through the fixed reference point is often called the angular momentum of the particle about this axis.

We define the torque (or turning moment) **N** about the same fixed point as

$$\boxed{\mathbf{N} \equiv \mathbf{r} \times \mathbf{F},} \tag{44}$$

where **F** is the force acting on the particle. The units of torque are dyne-cm. Now on differentiating (43) we have

$$\frac{d\mathbf{J}}{dt} = \frac{d}{dt}(\mathbf{r} \times \mathbf{p}) = \frac{d\mathbf{r}}{dt} \times \mathbf{p} + \mathbf{r} \times \frac{d\mathbf{p}}{dt}. \tag{45}$$

But

$$\frac{d\mathbf{r}}{dt} \times \mathbf{p} = \mathbf{v} \times M\mathbf{v} = 0, \tag{46}$$

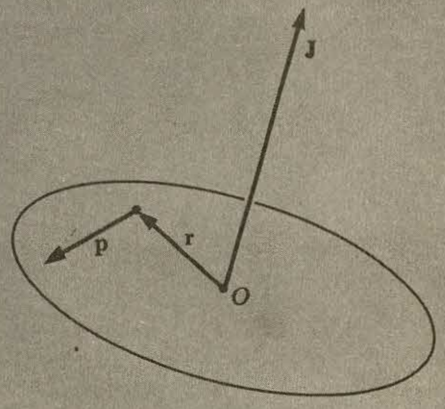
and by Newton's second law in an inertial reference frame,

$$\mathbf{r} \times \frac{d\mathbf{p}}{dt} = \mathbf{r} \times \mathbf{F} = \mathbf{N}. \tag{47}$$

Thus we have the important result

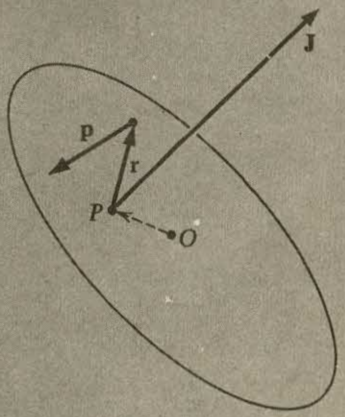
$$\boxed{\frac{d\mathbf{J}}{dt} = \mathbf{N};} \tag{48}$$

the time rate of change of angular momentum is equal to the torque.



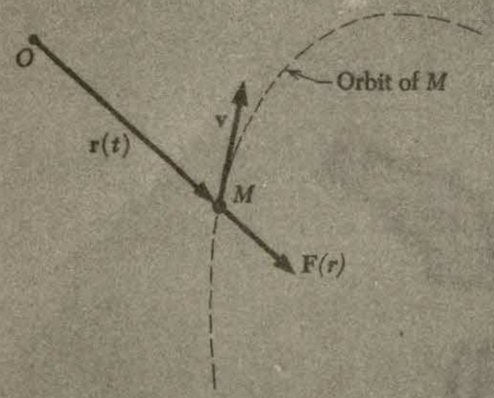
$\mathbf{J} = \mathbf{r} \times \mathbf{p}$

The angular momentum **J** *with respect to point O* is defined by the figure.



$\mathbf{J} = \mathbf{r} \times \mathbf{p}$

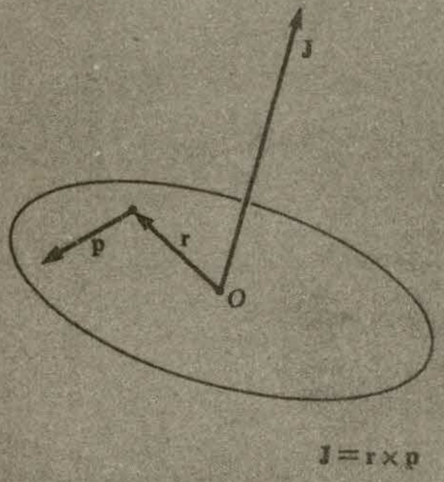
The angular momentum *with respect to another point P* is different even for the *same* particle with the *same* momentum **p**.



Particle M subjected to repulsive central force $\mathbf{F}(\mathbf{r})$ centered at O. Since the torque

$$\mathbf{N} = \mathbf{r} \times \mathbf{F} = 0,$$

we have $\mathbf{J} = \text{const.}$



Since $\mathbf{J}$ is a constant in central force motion, such motion is confined to a plane.

If the torque $\mathbf{N} = 0$, then $\mathbf{J} = \text{const.}$ *Angular momentum is constant in the absence of external torque*[†]; this is a statement of the *law of conservation of angular momentum.* Note that the law of conservation of angular momentum does not refer only to particles in closed orbits. It applies as well to open orbits and to collision processes.

Consider a particle subjected to a central force of the form

$$\mathbf{F} = \hat{\mathbf{r}} f(r). \tag{49}$$

A *central* force is one which is everywhere directed exactly toward (or away from) a particular point. The torque is

$$\mathbf{N} = \mathbf{r} \times \mathbf{F} = \mathbf{r} \times \hat{\mathbf{r}} f(r) = 0, \tag{50}$$

so that for central forces

$$\frac{d\mathbf{J}}{dt} = 0, \tag{51}$$

and the angular momentum is constant. The next section will show that this result can be viewed as a direct consequence of the invariance of the potential energy function $U(\mathbf{r})$ under rotations of the reference frame. First, however, we consider the extension of the torque equation to a system of N interacting particles.

The total angular momentum of a system of particles referred to an arbitrary fixed point in an inertial reference frame as origin is

$$\mathbf{J} = \sum_{n=1}^{N} M_n \mathbf{r}_n \times \mathbf{v}_n. \tag{52}$$

Just as for the single particle, the value of $\mathbf{J}$ depends on the point we choose for origin O. With $\mathbf{R}_{\text{c.m.}}$ as the vector from the origin to the position of the center of mass, we rewrite $\mathbf{J}$ in a convenient and important form as

$$\begin{aligned} \mathbf{J} = \sum_{n=1}^{N} M_n(\mathbf{r}_n - \mathbf{R}_{\text{c.m.}}) \times \mathbf{v}_n \\ + \sum_{n=1}^{N} M_n \mathbf{R}_{\text{c.m.}} \times \mathbf{v}_n = \mathbf{J}_{\text{c.m.}} + \mathbf{R}_{\text{c.m.}} \times \mathbf{P}, \end{aligned} \tag{53}$$

where $\mathbf{J}_{\text{c.m.}}$ is the angular momentum *about* the center of mass and $\mathbf{P} \equiv \Sigma M_n \mathbf{v}_n$ is the total linear momentum. The term $\mathbf{R}_{\text{c.m.}} \times \mathbf{P}$ is the angular momentum *of* the center of mass about the origin. This term depends on the choice of origin; the term $\mathbf{J}_{\text{c.m.}}$ does not. We sometimes find it useful for a *single particle* to call $\mathbf{J}_{\text{c.m.}}$ the *spin* angular momentum.

[†] An external torque is produced by an external force exerted on the body. An internal torque is associated with a force exerted on one part of the body by another part of the body.

Example. Internal torques. The interactions which may be present between the particles themselves give rise to internal torques. Show that *the sum of all internal torques is zero.*

The total torque is defined as

$$\mathbf{N} = \sum_{n=1}^{N} \mathbf{r}_n \times \mathbf{F}_n, \tag{54}$$

where for internal forces

$$\mathbf{F}_i = \sum_{j=1}^{N}{}' \mathbf{F}_{ij}, \tag{55}$$

the sum of the forces on particle i from all other particles j. (The prime on the summation sign means that the term $j = i$ is excluded.) Thus the internal torque is

$$\mathbf{N}_{\text{int}} = \sum_i \mathbf{r}_i \times \mathbf{F}_i = \sum_i \sum_j{}' \mathbf{r}_i \times \mathbf{F}_{ij}; \tag{56}$$

but by relabeling the dummy indices i and j, we have

$$\sum_i \sum_j{}' \mathbf{r}_i \times \mathbf{F}_{ij} \equiv \sum_j \sum_i{}' \mathbf{r}_j \times \mathbf{F}_{ji}, \tag{57}$$

so that

$$\mathbf{N}_{\text{int}} = \frac{1}{2} \sum_i \sum_j{}' (\mathbf{r}_i \times \mathbf{F}_{ij} + \mathbf{r}_j \times \mathbf{F}_{ji}). \tag{58}$$

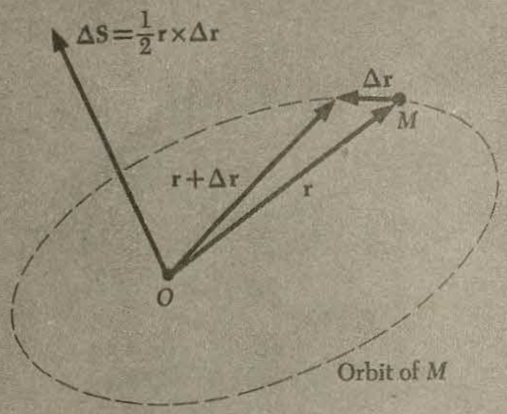
We now assume that the forces are Newtonian, which means we assume $\mathbf{F}_{ji} = -\mathbf{F}_{ij}$, whence

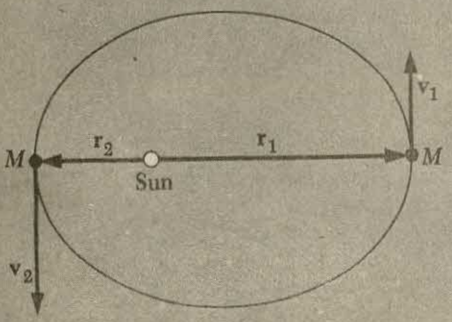
$$\mathbf{N}_{\text{int}} = \frac{1}{2} \sum_i \sum_j{}' (\mathbf{r}_i - \mathbf{r}_j) \times \mathbf{F}_{ij}. \tag{59}$$

For central forces $\mathbf{F}_{ij}$ is parallel to $\mathbf{r}_i - \mathbf{r}_j$. It is immediately obvious that $(\mathbf{r}_i - \mathbf{r}_j) \times \mathbf{F}_{ij} = 0$, and thus

$$\mathbf{N}_{\text{int}} = 0. \tag{60}$$

(The same result can be obtained for noncentral internal forces. The static noncentral forces of importance in physics can be simulated for most purposes by the forces from spatial arrays of particles, each particle giving rise to a central force.)





The planet M has constant angular momentum about the sun. Thus $Mr_2v_2 = Mr_1v_1$, where r_1 = greatest distance from sun and r_2 = smallest distance from sun. All the planetary orbits have much less eccentricity than shown here. The figure is exaggerated for clarity.

Assuming that $\mathbf{N}_{\text{int}} = 0$, then we have from (51), (53), and (60):

$$\boxed{\begin{aligned} \frac{d}{dt}\mathbf{J}_{\text{total}} &= \mathbf{N}_{\text{external}}; \\ \mathbf{J}_{\text{total}} &= \mathbf{J}_{\text{c.m.}} + \mathbf{R}_{\text{c.m.}} \times \mathbf{P}. \end{aligned}} \qquad \begin{aligned} (61) \\ (62) \end{aligned}$$

Here $\mathbf{J}_{\text{c.m.}}$ is the angular momentum about the center of mass, whereas $\mathbf{R}_{\text{c.m.}} \times \mathbf{P}$ is the angular momentum of the center of mass about the arbitrary origin. It is usually a very good idea to choose the origin at the center of mass. Then (61) may be written as

$$\frac{d}{dt}\mathbf{J}_{\text{c.m.}} = \mathbf{N}_{\text{ext}}. \tag{63}$$

If no external forces act, then $\mathbf{N}_{\text{ext}} = 0$ and $\mathbf{J}_{\text{c.m.}}$ is constant.

We saw that the motion of the center of mass is determined by the total external force acting on the body. We see now that the rotation about the center of mass is determined by the total external torque.

The geometric meaning of the angular momentum of a particle in an orbit enclosing the origin is suggested by the accompanying figure. The vector area $\Delta\mathbf{S}$ of the triangle is given by

$$\Delta\mathbf{S} = \frac{1}{2}\mathbf{r} \times \Delta\mathbf{r}. \tag{64}$$

Then

$$\frac{d\mathbf{S}}{dt} = \frac{1}{2}\mathbf{r} \times \mathbf{v} = \frac{1}{2M}\mathbf{r} \times \mathbf{p} = \frac{1}{2M}\mathbf{J}. \tag{65}$$

We have seen that, with suitable choice of the origin, $\mathbf{J}$ = constant for central forces.

If in a planetary problem the origin is taken at the sun, then the angular momentum of a planet is constant, apart from disturbances (perturbations) by other planets. For central forces we see from (64) and (65) that

1. The orbit lies in a plane.
2. The rate of sweeping out of area is a constant—this is one of the three Kepler laws (discussed in Chap. 9).

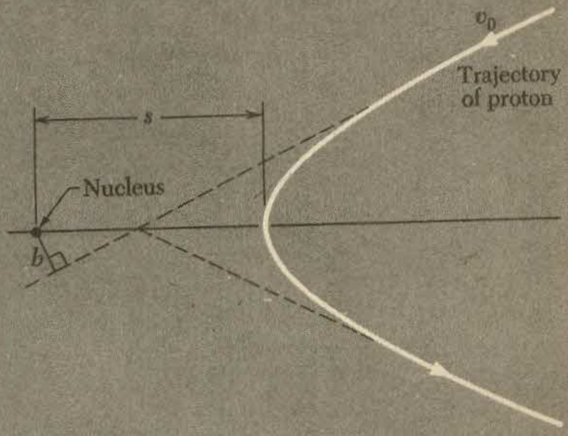
The first result follows because $\mathbf{r}$ and $\Delta\mathbf{r}$ are in a plane perpendicular to $\mathbf{J}$, and $\mathbf{J}$ is constant in magnitude and direction in a central field.

For a particle moving in a circle the velocity **v** is perpendicular to **r**, so that

$$J = Mvr = M\omega r^2. \tag{66}$$

The planets move in elliptical orbits about the sun at a focus. In order to conserve angular momentum each planet must move faster at the point of closest approach than at the furthest point. This result follows because at these points **r** is perpendicular to **v** and the angular momentum at these points is Mvr. By the conservation of angular momentum the values of Mvr at these points must be equal, so that the shorter r is associated with the larger v.

Example. Proton scattering by heavy nucleus. A proton approaches a very massive nucleus of charge Ze. At infinite separation the energy of the proton is $\frac{1}{2}M_pv_0^2$. A *linear extrapolation* to small separations of the trajectory at large separations would pass at a minimum distance b from the heavy nucleus, as in the figure. This distance is called the impact parameter.



Motion of proton in Coulomb field of a heavy nucleus. The trajectory is a hyperbola (Chap. 9). The distance of closest approach is s. The impact parameter b is the normal distance from the nucleus to the linear extrapolation of the initial portion of the trajectory.

What is the distance of closest approach for the actual orbit? Take the mass of the heavy nucleus to be infinite, so that its recoil energy ~~may~~ be neglected. (The complete scattering problem is given as the historical note to Chap. 15.)

The initial angular momentum of the proton taken about the heavy nucleus is M_pv_0b, where v_0 is the initial velocity of the proton. At the distance of closest approach, denoted by s, the angular momentum is M_pv_ss, where v_s is the velocity at this point. The force is central, so that angular momentum is conserved and thus

$$M_pv_0b = M_pv_ss; \qquad v_s = \frac{v_0b}{s}. \tag{66a}$$

Note that we have neglected the angular momentum transfer to the heavy nucleus.

The energy of the proton is also conserved in the collision. The initial energy is all kinetic and is $\frac{1}{2}M_pv_0^2$. The energy at the point of closest approach is

$$\frac{1}{2}M_pv_s^2 + \frac{Ze^2}{s}, \tag{66b}$$

where the first term is the kinetic energy and the second term is the potential energy. Thus the law of conservation of energy tells us that

$$\frac{1}{2}M_p v_s^2 + \frac{Ze^2}{s} = \frac{1}{2}M_p v_0^2, \tag{66c}$$

or, using our earlier result,

$$\frac{Ze^2}{s} = \frac{1}{2}M_p v_0^2\left[1 - \left(\frac{b}{s}\right)^2\right]. \tag{66d}$$

This equation may be solved for s. Note that the conservation laws have told us quite a lot about the collision process.

Rotational Invariance. Conservation of angular momentum is a consequence of the invariance of the potential energy under rotation of the reference frame (or of the system). If there is an external torque, we will in general do work against this torque in rotating the system. If we do work, the potential energy must change. If the potential energy U is unchanged by the rotation, there is no external torque. Zero external torque means that the angular momentum is conserved.

The argument can be pursued analytically in exact analogy to the earlier argument, Eqs. (1) to (7), for the conservation of linear momentum. Let $\Omega\mathbf{r}$ represent the vector which is obtained† by rotating $\mathbf{r}$ through an arbitrary angle θ about an arbitrary axis. The length of $\Omega\mathbf{r}$ is the same as that of $\mathbf{r}$. We assert that the *conservation of angular momentum follows from the rotational invariance defined by*

$$U(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N) = U(\Omega\mathbf{r}_1, \Omega\mathbf{r}_2, \ldots, \Omega\mathbf{r}_N). \tag{67}$$

What does this relation imply about the dependence of U upon its argument for two particles? Consider the special form

$$U = U(\mathbf{r}_1 - \mathbf{r}_2). \tag{68}$$

The replacement of $\mathbf{r}$ by the rotated vector $\Omega\mathbf{r}$ changes the direction but not the magnitude of $\mathbf{r}_1 - \mathbf{r}_2$. Thus if U is to be invariant, it can depend only on the magnitude $|\mathbf{r}_1 - \mathbf{r}_2|$ of the distance between the two particles. That is,

$$U(\mathbf{r}_1,\mathbf{r}_2) \equiv U(|\mathbf{r}_1 - \mathbf{r}_2|). \tag{69}$$

---

† Observe that here we rotate $\mathbf{r}$ and not the reference frame; we can do either, but the present argument may seem simpler if $\mathbf{r}$ is rotated. The quantity Ω is not a simple number; it is called an *operator*. We shall consider operators in Vol. IV.

For this potential the force $\mathbf{F}_{12} = -\mathbf{F}_{21}$ is automatically directed along the line $\mathbf{r}_2 - \mathbf{r}_1$. The force is central and by (50) the torque vanishes.

For N particles the rotational invariance of the potential is assured if U depends only on the magnitude of the separation between the several particles.

The potential seen by an individual electron or ion in a crystal is not rotationally invariant because the electric field due to the other ions in a crystal is highly nonuniform or inhomogeneous. Therefore we do not in general expect to find a conservation law for the angular momentum of the electronic shells of an ion in a crystal, even though the angular momentum is conserved for the same ion considered in free space. The nonconservation of electronic angular momentum of ions in crystals has been observed in studies of paramagnetic ions in crystals, and the effect is called quenching of the orbital angular momentum.

The angular momentum $\mathbf{J}$ of the earth is constant with respect to the sun as origin because $\mathbf{r} \times \mathbf{F} = 0$ for each mass point in the earth, where $\mathbf{F}$ is the gravitational force acting between the sun and the mass point.

Example. Angular acceleration accompanying contraction. A particle of mass M is attached to a string; the particle rotates with velocity v_0 when the length of the string is r_0. How much work is done in shortening the string to r?

The force on the particle due to the string is radial, so that *the torque is zero* as the string is shortened. Therefore the angular momentum must remain constant as the string is shortened:

$$Mv_0r_0 = Mvr. \tag{70}$$

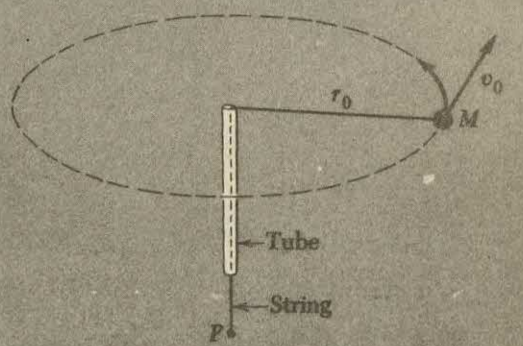
The kinetic energy at r_0 is $\frac{1}{2}Mv_0^2$; at r it has been increased to

$$\frac{1}{2}Mv^2 = \frac{1}{2}Mv_0^2\left(\frac{r_0}{r}\right)^2, \tag{71}$$

because $v = v_0r_0/r$ from (70). It follows that the work W done from outside in shortening the string from r_0 to r is

$$W = \frac{1}{2}Mv_0^2\left[\left(\frac{r_0}{r}\right)^2 - 1\right]. \tag{72}$$

We see that the angular momentum acts on the radial motion as an effective repulsive potential energy: we have to do *extra* work on the particle on bringing it in from large distances to



Mass M describes circular motion of radius r_0 and velocity v_0. It is connected to a string which passes through a tube. The distance r_0 can be shortened by pulling on the string at P.

small distances *if* we require that the angular momentum be conserved in the process.

Compare this behavior with that of a particle rotating on a string that is freely winding up on a smooth fixed peg of finite diameter. Why is the kinetic energy now constant as the string winds up?

Shape of the Galaxy. The result of the preceding example has a probable bearing on the shape of the Galaxy. Consider a very large mass M of gas endowed initially with some angular momentum.† The gas contracts under its gravitational interaction. As the volume occupied by the gas gets smaller, the conservation of angular momentum requires an increase in the angular velocity. But we have just seen that work is needed to produce the increase in angular velocity. Where does the kinetic energy come from? It can only come from the gravitational energy of the gas.

A particle of mass M_1 in the outer regions of the galaxy will have a gravitational potential energy due to the interaction of the particle with the galaxy of the order of magnitude of

$$-\frac{GM_1M}{r}, \tag{73}$$

where r is the distance from the center of the galaxy and M is the mass of the galaxy. The effective potential energy associated with the angular momentum depends on r as in (72).

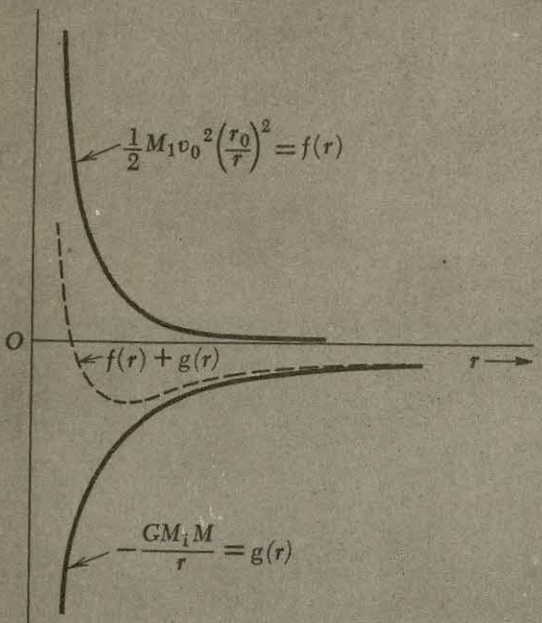
The sum of (72) and (73) is an extremum (this is a necessary condition for equilibrium) when

$$\frac{d}{dr}\left[-\frac{GM_1M}{r}+\frac{1}{2}M_1v_0^2\left(\frac{r_0}{r}\right)^2\right]=0, \tag{74}$$

or when

$$\frac{GM_1M}{r^2}-M_1v_0^2\frac{r_0^2}{r^3}=0; \qquad r=\frac{v_0^2r_0^2}{GM}=\frac{v^2r^2}{GM}. \tag{75}$$

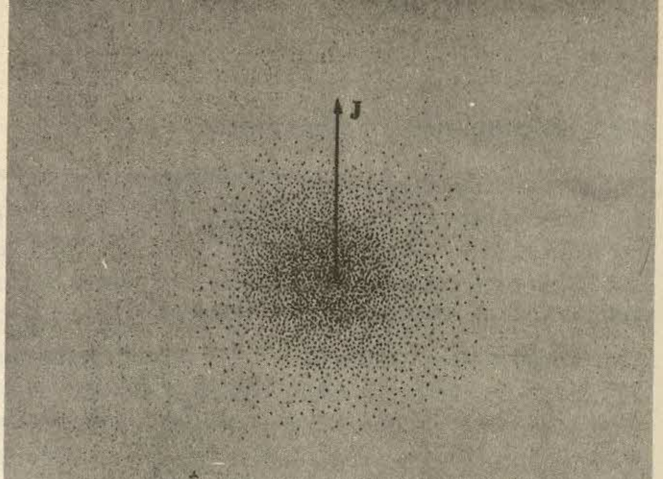
At smaller values of r than that given by (75) the gravitational energy is not strong enough to continue to drive the contraction. Equation (75) is identical with the statement that when $r = r_0$ the attractive force GM_1M/r^2 is equal to the centrifugal force M_1v^2/r.



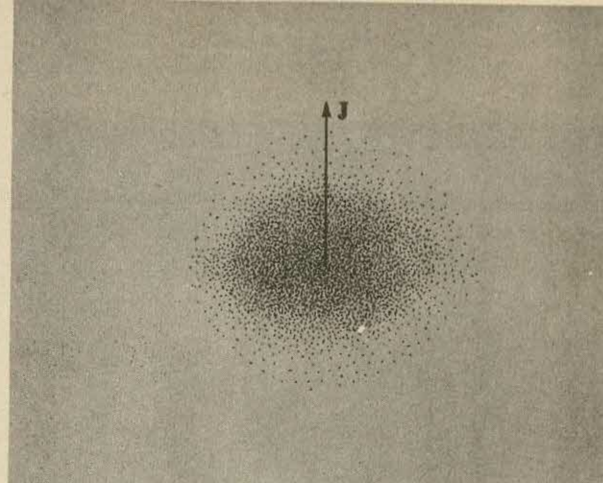
Contraction of the galaxy in the plane normal to **J** is limited because the centrifugal "potential energy" $f(r)$ increases so rapidly as $r \to 0$. Thus $f(r) + g(r)$ has a minimum at finite value of r, as shown.

---

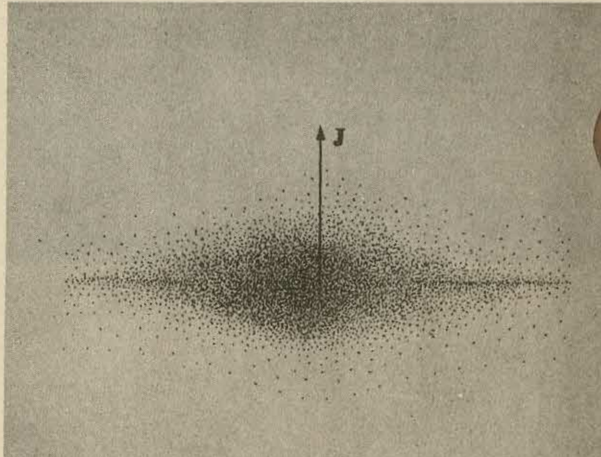
† It is not possible in the present state of knowledge to say where the gas came from in the first place, or why a given mass of gas should have an angular momentum. Masses without angular momentum will condense as spheres.



Originally a cloud of gas, . . .



the galaxy begins to flatten, . . .



and finally assumes a pancake shape with a more or less spherical central core.

But the cloud of gas or of stars is able to collapse in the direction parallel to the axis of the total angular momentum without changing the value of the angular momentum. The contraction is driven by gravitational attraction; the energy gain in contraction must be dissipated in some way, and this is believed to occur by radiation. The cloud is therefore able to collapse rather completely in the direction parallel to **J**, but the contraction in the equatorial plane is restricted. This model of galactic evolution is due to Hubble; some recent workers believe it may be oversimplified.

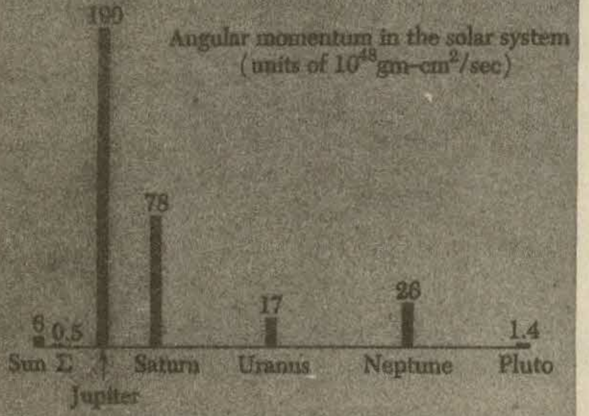
The diameter of our galaxy is of the order of 3×10^4 parsecs, or 10^{23} cm. (1 parsec $= 3.084 \times 10^{18}$ cm.) The thickness of the galaxy in the neighborhood of the sun depends somewhat on how the thickness is defined, but the vast majority of stars cluster about the median plane in a thickness of several hundred parsecs. Thus the galaxy is greatly flattened. The mass of the galaxy is put at about 2×10^{11} times the mass of the sun, or

$$(2 \times 10^{11})(2 \times 10^{33}) \approx 4 \times 10^{44} \text{ gm.}$$

An estimate of the mass may be made from (75) by substituting the known values of v and r for the sun. The sun lies toward the outer edge of the galaxy at about 10^4 parsecs $\approx 3 \times 10^{22}$ cm from the axis of the galaxy. The orbital velocity of the sun about the center of the galaxy is approximately 3×10^7 cm/sec, so that from (75) we derive as an estimate of the mass of the galaxy

$$M = \frac{v^2 r}{G} \approx \frac{(10^{15})(3 \times 10^{22})}{7 \times 10^{-8}} \approx 4 \times 10^{44} \text{ gm.} \qquad (76)$$

We have neglected the effect of the mass which lies further from the center of the galaxy than the sun.



Distribution of angular momentum in the solar system, about the center of the sun. The symbol Σ denotes the sum of the four planets Mercury, Venus, Earth, and Mars. Note the relatively small contribution of the rotation of the sun about its own axis.

Angular Momentum of the Solar System. The accompanying figure shows the angular momentum of the several components of the solar system. Let us make an estimate for ourselves of one of the values given, as a check. Consider the planet Neptune, whose orbit is very closely circular. The mean distance of Neptune from the sun is given in reference works as 2.8×10^9 miles $\approx 5 \times 10^9$ km $\approx 5 \times 10^{14}$ cm. The period of revolution of Neptune about the sun is 165 years $\approx 5 \times 10^9$ sec. The mass of Neptune is about 1×10^{29} gm. The angular momentum of Neptune about the sun is

$$J = Mvr = M\frac{2\pi r}{T}r \approx \frac{(10^{29})(6)(25 \times 10^{28})}{5 \times 10^9}$$
$$\approx 30 \times 10^{48} \text{ gm-cm}^2/\text{sec}, \tag{77}$$

in approximate agreement with the value 26×10^{48} gm-cm^2/sec indicated on the figure. The direction of **J** is roughly the same for all the major planets.

The angular momentum of Neptune *about its own center of mass* is much smaller. The angular momentum of a rotating uniform sphere is of the order of MvR, where v is the surface velocity from the rotation and R the radius. Actually, because the mass of a sphere is not concentrated at a distance R from the axis, but is distributed, this result must be reduced for uniform distribution by a numerical factor which in Chap. 8 is found to be $\frac{2}{5}$. Thus

$$J_{\text{c.m.}} = \frac{2}{5}\frac{2\pi MR^2}{T}, \tag{78}$$

where $T \equiv 2\pi R/v$ denotes the period of rotation of the planet about its own axis. For Neptune $T \approx 16$ hr $\approx 6 \times 10^4$ sec and $r \approx 1.5 \times 10^4$ mi $\approx 2.4 \times 10^9$ cm, whence

$$J_{\text{c.m.}} \approx \frac{(0.4)(6)(10^{29})(6 \times 10^{18})}{6 \times 10^4} \approx 2 \times 10^{43} \text{ gm-cm}^2/\text{sec}, \tag{79}$$

which is negligible compared to the orbital angular momentum about the sun, as given in (77).

A similar estimate of $J_{\text{c.m.}}$ for the sun gives 6×10^{48} gm-cm^2/sec. The rotation of the sun about an axis through its center accounts for only about 2 percent of the total angular momentum in the solar system. A typical hotter star may carry about 100 times as much angular momentum as the sun. It thus appears that the formation of a planetary system is an effective mechanism for carrying off angular momentum from a cooling star. If every star forms a planetary system in passing through the stage of its history similar to the sun, then there may be over 10^{10} stars with planets in our galaxy.

Intrinsic Angular Momentum of Elementary Particles. It is known from experiments discussed in detail in Vol. IV that the elementary particles have an intrinsic angular momentum $J_{\text{c.m.}}$. The intrinsic angular momentum is usually called the *spin angular momentum.* The spin angular momentum of an elementary particle is denoted by S and is usually measured in units of

$$\hbar = \frac{\text{Planck's constant}}{2\pi} = 1.0542 \times 10^{-27} \text{ erg-sec.}$$

We note that the dimensions of $\hbar$ are the same as the dimensions of Mvr. Values of S for several elementary particles are tabulated below:

Particle	*Spin angular momentum, S*
Electron	$\frac{1}{2}$
Photon	1
Nucleon (proton or neutron)	$\frac{1}{2}$
Neutrino	$\frac{1}{2}$
$\mu^{\pm}$ meson	$\frac{1}{2}$
$\pi^{\pm}$, π^0 meson	0
Λ^0 (hyperon)	$\frac{1}{2}$
$K^{\pm}$, K^0 meson	0

Further reading

PSSC, chap. 23

Film list

"Vorticity" A. Shapiro (ESI film). Gives some beautiful demonstrations of fluid dynamics which illustrate line integrals, curls, and angular momentum.

Problems

1. *Angular momentum of a satellite.* (*a*) What is the angular momentum (referred to the center of the orbit) of a satellite of mass M_s which moves in a circular orbit of radius r? The result is to be expressed in terms only of r, G, M_s, M_e (the mass of the earth).

Ans. $J = (GM_eM_s^2r)^{1/2}$.

(*b*) For $M_s = 100$ kg, what is the numerical value (in CGS units) of the angular momentum of an orbit for which the radius is twice the radius of the earth?

2. *Frictional effects on satellite motion.* (*a*) What is the effect of atmospheric friction on the motion of a satellite in a circular (or nearly circular) orbit? Why does friction increase the satellite velocity?

(*b*) Does friction increase or decrease the angular momentum of the satellite, measured with respect to the center of the earth? Why?

3. *Energy–angular momentum relation for a satellite.* Express in terms of the angular momentum J the kinetic, potential, and total energy of a satellite of mass M in a circular orbit of radius r.

Ans. $K = J^2/2Mr^2$; $U = -J^2/Mr^2$; $E = -J^2/2Mr^2$.

4. *Electron bound to a proton.* An electron moves about a proton in a circular orbit of radius $0.5\ \text{Å} \equiv 0.5 \times 10^{-8}$ cm.

(*a*) What is the orbital angular momentum of the electron about the proton? *Ans.* 1×10^{-27} erg-sec.

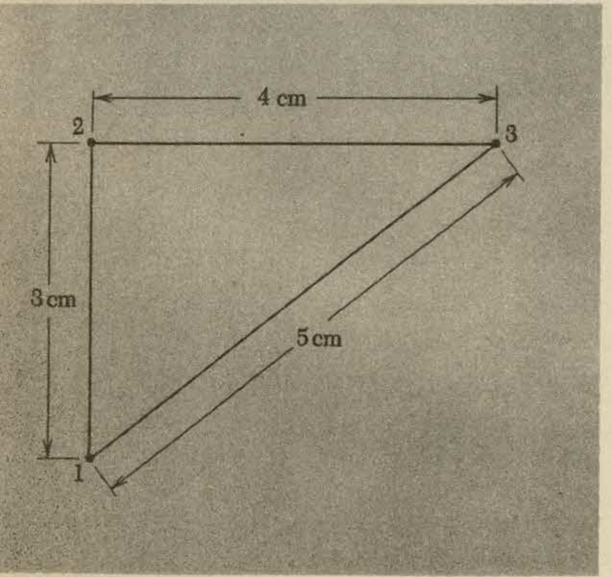
(*b*) What is the total energy, expressed in ergs and in electron volts?

5. *Internal torques sum to zero.* Consider the isolated system of three particles 1, 2, 3, (shown in the diagram) interacting with *central* forces $F_{12} = 1$ dyne, $F_{13} = 0.6$ dyne, $F_{23} = 0.75$ dyne, where F_{ij} denotes the force on particle i when it interacts with particle j.

(*a*) Show explicitly, writing all forces in component form in an appropriate coordinate system, that the sum of the internal torques is zero:

$$\mathbf{N}_{\text{int}} = \sum_{i=1}^{3} \sum_{j\neq 1}^{3} \mathbf{r}_i \times \mathbf{F}_{ij} = 0.$$

(*b*) Show that the same result is obtained by interchanging the dummy indices and summing $\mathbf{r}_j \times \mathbf{F}_{ji}$.



6. *Forces on ladder.* A ladder of mass 20 kg and length 10 m rests against a slippery vertical wall, at an angle of 30° with the vertical. The ladder, of uniform construction, is prevented from slipping by friction with the ground. What is the magnitude in dynes of the force exerted by the ladder on the wall? (*Hint:* Use the fact that the torques must sum to zero for a ladder at rest.) *Ans.* 5.6×10^6 dynes.

7. *Rotational kinetic energy.* What is the rotational kinetic energy (in ergs) of a thin circular hoop, of radius 1 m and linear density 1 gm/cm, rotating at 100 cycles/sec about an axis through the center of the hoop (and normal to the plane of the hoop)?

8. *Angular momentum of moon.* Compare the value of the angular momentum of the moon in its orbit about the earth with the value of the angular momentum of rotation $\frac{2}{5}M\omega R_M^2$ of the moon about its own axis.

9. *Angular momentum in near collision of two particles.* A neutron of energy 1 Mev passes a proton at such a distance that the angular momentum of the neutron relative to the proton approximately equals 10^{-26} erg-sec. What is the distance of closest approach? (We neglect the energy of interaction between the two particles.) *Ans.* 4×10^{-12} cm.

10. *Angular momentum of linear motion.* A particle of mass M travels with velocity $\mathbf{v}_1 = v\hat{\mathbf{x}}$ along the line $y = y_1$. At $t = 0$ the particle is at the point $(0,y_1)$.

(*a*) Calculate the angular momentum of the particle about the origin.
Ans. $-Mvy_1\hat{\mathbf{z}}$.

(*b*) Calculate the total angular momentum about the point $(0,y_2)$, where $y_2 < y_1$.

11. *Particle-dumbbell collision.* Two equal masses M are connected by a rigid rod of negligible mass and of length a. The center of mass of this dumbbell-like system is stationary in gravity-free space, and the system rotates about the center of mass with angular velocity $\boldsymbol{\omega}$. One of the rotating masses strikes a third stationary mass M, which sticks to it.

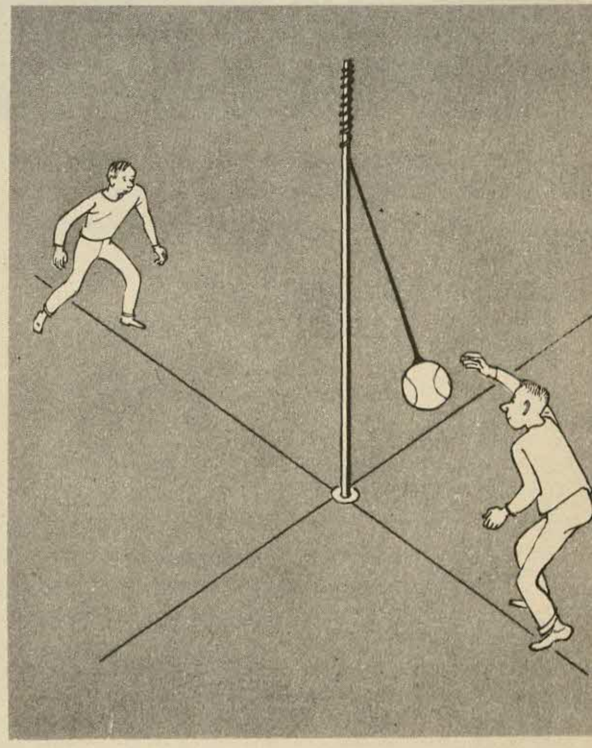
(*a*) Locate the center of mass of the three-particle system at the instant prior to collision. What is the velocity of the center of mass?

(*b*) What is the angular momentum of the three-mass system about the center of mass at the instant prior to collision? At the instant following collision?

(*c*) What is the angular velocity of the system about the center of mass after the collision?

(*d*) What are the initial and final kinetic energies?

12. *Angular momentum of tetherball.* The object of the game tetherball is to hit the ball hard enough and fast enough to wind its tether cord in one direction about the vertical post to which it is tied before the opposing player can wind it in the opposite direction. The game is exciting, and the dynamics of the ball's motion are complicated. Let us examine a simple type of motion in which the ball moves in a horizontal plane in a spiral of decreasing radius as the cord winds round the post after a single blow which gives the ball an initial speed v_0. The length of the cord is l and the radius of the post is $a \ll l$ (refer to the diagram).



(*a*) What is the instantaneous center of revolution?

(*b*) Is there a torque about the axis through the center of the post? Is angular momentum conserved?

(*c*) Assume that kinetic energy is conserved and calculate the speed as a function of time.

(*d*) What is the angular velocity after the ball has made five complete revolutions? *Ans.* $\omega = (l - 10\pi a)v_0/[a^2 + (l - 10\pi a)^2]$.

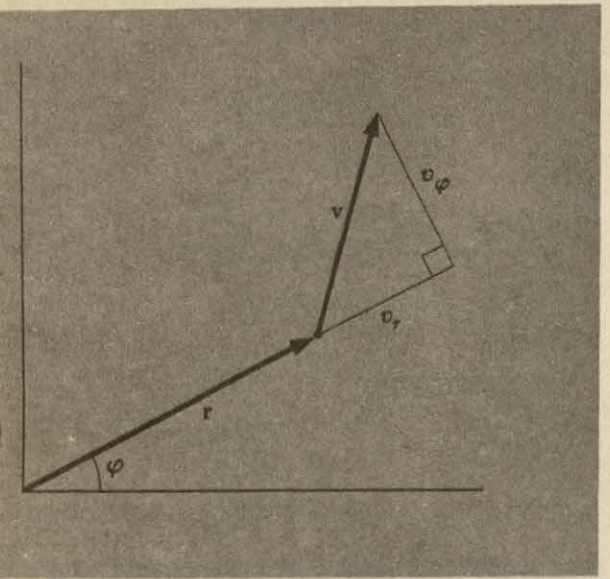
13. *Effective centrifugal potential energy.* It is convenient to use plane polar coordinates r, φ for motion in a plane perpendicular to an axis of rotation.

(*a*) Show that the velocity in such a coordinate system may be written

$$\mathbf{v} = v_r\hat{\mathbf{r}} + v_\varphi\hat{\boldsymbol{\varphi}},$$

where v_r is just dr/dt, the rate of change of the length r, and

$$v_\varphi = r\,d\varphi/dt.$$



(*b*) Show that the kinetic energy of a particle in this coordinate system is

$$K = \tfrac{1}{2}M(\dot{r}^2 + \omega^2 r^2),$$

where $\omega = d\varphi/dt$.

(*c*) Show that the total energy is

$$E = U(r) + \frac{1}{2}M\dot{r}^2 + \frac{J^2}{2Mr^2},$$

where J is the angular momentum of the particle about the fixed axis normal to the plane of the motion. *Hint:* Recall Eq. (72).

(*d*) Because the force is central there is no torque on the particle and J is a constant of the motion. The term $J^2/2Mr^2$ is sometimes called the centrifugal potential energy. Show that the centrifugal potential energy represents an outward radial force J^2/Mr^3.

(*e*) If $U(r) = \tfrac{1}{2}Cr^2$, show that $U(r)$ represents an inward radial force $-Cr$.

(*f*) Show from (*d*) and (*e*) that the balance of these forces is equivalent to the condition $\omega^2 = C/M$.

14. *Angular momentum–torque relation.* If **J** and **N** are referred to a center of mass as origin, show that $d\mathbf{J}/dt = \mathbf{N}$ even if the center of mass has a variable velocity $\mathbf{v}(t)$ referred to some inertial frame.

Advanced Topic: Meteoroid's encounter with the atmosphere

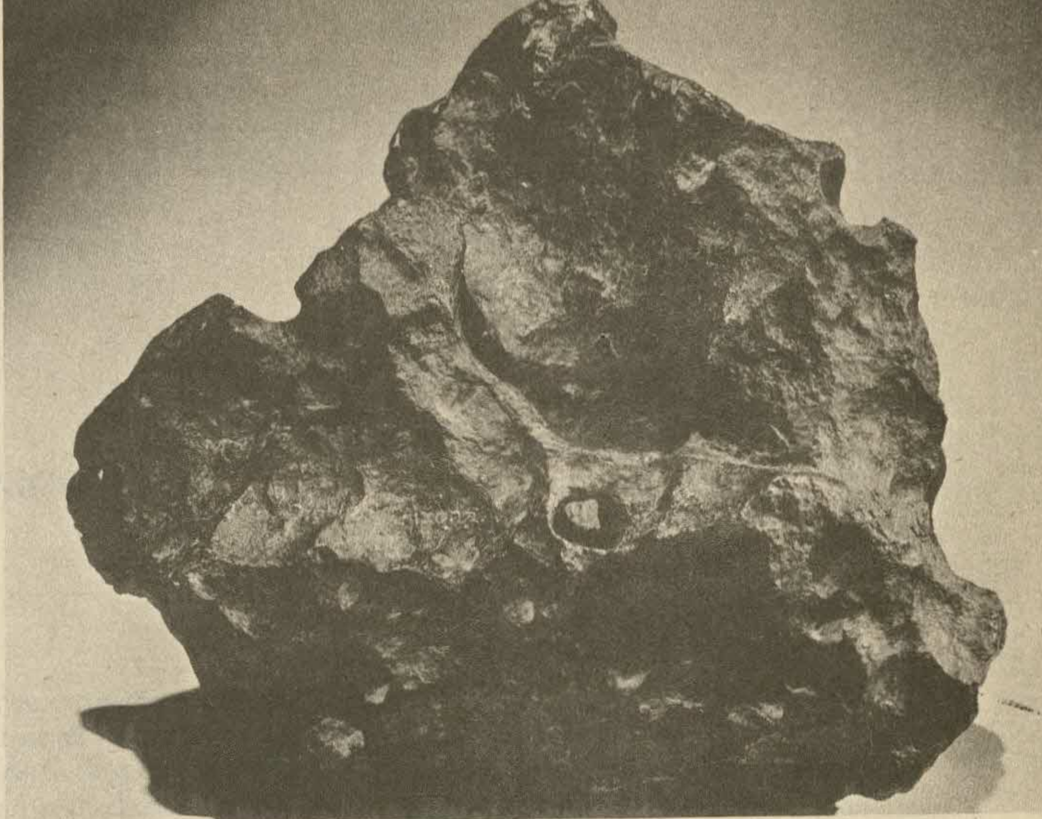
Meteoroids, which are small bodies in interplanetary space moving in closed orbits around the sun, occasionally collide with the earth, producing visible meteors. A meteoroid decelerates when it collides with the outer atmosphere of the earth because of the momentum exchange between the meteoroid and the air molecules with which it collides. When the average distance (mean free path) an air molecule travels between collisions with another molecule is large compared to the linear dimensions of the meteoroid, the deceleration problem may be thought of in terms of individual collisions of the air molecules with the meteoroid, rather than as a problem in classical hydrodynamics.

(*a*) Derive an expression for the deceleration in terms of the mass M, velocity v, and effective cross-sectional area† S of the meteoroid, on the rough assumption that the initial molecular velocity is negligible in comparison with v. This is a reasonable assumption because otherwise the air molecules would escape from the atmosphere. Show that

$$\frac{dv}{dt} = -\frac{\Gamma S \rho v^2}{M},$$

---

† When we say that the meteoroid has an effective cross-sectional area S, we mean that, when the meteoroid moves a distance x, it strikes a number of molecules equal to the number of molecules in the volume Sx.



A 95-pound piece of the Canyon Diablo meteorite. The piece is roughly one foot in diameter. (*Brookhaven National Laboratory photograph*)

where ρ is the density of the atmosphere and Γ is a numerical constant of the order of unity. The value of Γ will depend on whether you assume that the collisions are elastic or inelastic. Assume that the collisions are inelastic in that the molecules stick to the meteoroid. If you wish, treat the meteoroid as a cube traveling with a face perpendicular to the motion.

(*b*) The ablation (loss of mass) of the meteoroid can occur through fragmentation or vaporization. Using the principle of conservation of energy, derive an expression for the rate of mass loss in terms of the velocity, mass, and density of the meteoroid and the density of the air. Use the symbol ζ (zeta) for the constant giving the amount of energy required to ablate one gram of meteoroidal material by whatever processes occur and the symbol Λ for the constant giving the efficiency of exchange of energy between the air molecules and the meteoroid. Neglect here the kinetic energy lost by deceleration. The approximate result is

$$\frac{dM}{dt} = -\frac{\Lambda M^{2/3}\rho v^3}{2\zeta\rho_m{}^{2/3}},$$

where ρ_m is the density of the meteoroid, and where we have assumed that $S \equiv (M/\rho_m)^{2/3}$, which is not exact.

(*c*) Estimate from (*a*) the order of magnitude of dv/dt, for plausible values of the relevant quantities. Use ρ as estimated at an altitude of 100 km, which is 8×10^{-10} gm/cm^3. At this altitude the mean free path of a molecule of N_2 is about 10 cm. These data may be found in standard handbooks.

Chapter 7

Simple pendulum 192
Example: Nonlinear effects 197
Mass on a spring 200
Example: Average kinetic and potential energy 201
LC circuit 204
Friction 205
Damped harmonic oscillator 207
Example: Power dissipation 209
Quality factor Q 211
Driven harmonic oscillator 211
Low driving frequency, $\omega \ll \omega_0$ 214
Resonance response, $\omega = \omega_0$ 216
High driving frequency, $\omega \gg \omega_0$ 217
Power absorption 217
Example: Numerical example of a harmonic oscillator problem 218

Superposition principle 219

Further reading 220

Film list 220

Problems 220

Advanced Topic 1: Exact solution, problem of the simple pendulum 225

Advanced Topic 2: Anharmonic oscillator 227

Advanced Topic 3: Modulation of the oscillator parameters (parametric amplification) 229

Mathematical Note: Complex numbers and the driven harmonic oscillator 231

Harmonic Oscillator

The harmonic oscillator is an exceptionally important example of periodic motion because it serves as an exact or as an approximate model for many problems in classical and quantum physics. The classical systems which are realizations of the harmonic oscillator include any stable system when slightly displaced from equilibrium, such as

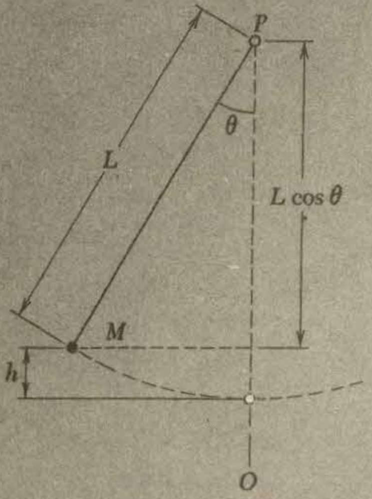
1. A simple pendulum, in the limit of small angle of oscillation.
2. A mass on a spring, in the limit of small amplitude of oscillation.
3. An electrical circuit composed of an inductance and a capacitance, for currents or voltages low enough for the circuit elements to be linear.

An electrical or mechanical circuit element is said to be linear if the response is directly proportional to the driving force. Most phenomena (but not all the interesting ones) in physics are linear if the range is taken to be small enough, just as most curves you encounter may be considered to be straight lines, for a sufficiently small range of values.

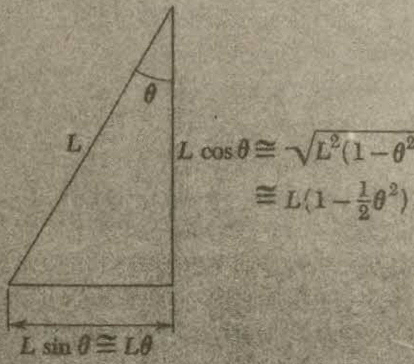
The most important properties of the harmonic oscillator are the following:

1. The frequency of the motion is independent of the amplitude of the oscillation.
2. The effects of several driving forces may be superposed linearly.

In this chapter we shall treat these properties of the harmonic oscillator. We shall consider both free and forced motion, as well as the effects of friction and of small anharmonic or nonlinear interactions added to the system. It is important to know what happens when the system can no longer be treated as linear.



The simple pendulum consists of a point mass M on the end of a massless rod L. The pendulum rotates about an axis through P which is perpendicular to the paper. Line OP is vertical.


θ
L
$L\cos\theta \cong \sqrt{L^2(1-\theta^2)}$
$\cong L(1-\frac{1}{2}\theta^2)$
$L\sin\theta \cong L\theta$

The Pythagorean theorem together with the binomial expansion reveals why $\cos\theta \cong 1 - \frac{1}{2}\theta^2$ for $\theta \ll 1$ radian.

Simple pendulum. The simple pendulum consists of a point mass M at the lower end of a massless rod of length L, pivoted freely at its upper end. We want to solve for the natural frequency of oscillation of the pendulum. The easiest way of setting up this problem is by writing the appropriate form of $\mathbf{F} = M\mathbf{a}$. This the student may do as Prob. 6. But it is very instructive to follow through a hammer-and-tongs approach, starting from the law of conservation of energy. Yet another approach, starting from the angular momentum, is given in Eqs. (18) to (22). The deflection is measured by the angle θ the rod makes with the vertical.

When the rod is deflected through the angle θ, the lower end of the rod is raised by the distance

$$h = L - L\cos\theta, \tag{1}$$

as is seen by reference to the figure. The potential energy of the mass M in the gravitational field of the earth is

$$U(h) = Mgh, \tag{2}$$

referred to the undeflected (vertical) position as the zero of potential energy. On substituting (1) in (2) we have

$$U(\theta) = MgL(1 - \cos\theta); \qquad U(0) = 0. \tag{3}$$

The kinetic energy of the pendulum is

$$K = \tfrac{1}{2}Mv^2 = \tfrac{1}{2}ML^2\dot{\theta}^2, \tag{4}$$

where $v = L\dot{\theta}$ relates the velocity and the rate of change of the angle of deflection. The total energy is

$$E = K.E. + P.E. = K + U = \tfrac{1}{2}ML^2\dot{\theta}^2 + MgL(1 - \cos\theta). \tag{5}$$

By the law of conservation of energy we know that this sum must be constant. We use this fact to obtain simply a solution for the frequency of the motion. Many readers will prefer the shorter approach of Eqs. (18) to (22) below. Now

$$\cos\theta \cong 1 - \tfrac{1}{2}\theta^2; \tag{6}$$

thus for $\theta \ll 1$ radian we may approximate the energy (5) by

$$E = \tfrac{1}{2}ML^2\,\dot{\theta}^2 + \tfrac{1}{2}MgL\,\theta^2. \tag{7}$$

Solving (7) for $\dot{\theta}$ we find

$$\frac{d\theta}{dt} = \left(\frac{2E - MgL\,\theta^2}{ML^2}\right)^{1/2} = \left(\frac{g}{L}\right)^{1/2}\left(\frac{2E}{MgL} - \theta^2\right)^{1/2}. \tag{8}$$

We denote the turning points of the motion by θ_0 and $-\theta_0$; the amplitude of the oscillation is θ_0. At these points the pendulum is momentarily at rest and the kinetic energy is zero. From (7) with $\dot{\theta} = 0$ we have

$$E = \frac{1}{2}MgL\,\theta_0{}^2; \qquad \theta_0{}^2 = \frac{2E}{MgL}. \tag{9}$$

Thus we may rewrite (8) as

$$\frac{d\theta}{dt} = \left(\frac{g}{L}\right)^{1/2}(\theta_0{}^2 - \theta^2)^{1/2}, \tag{10}$$

or

$$\frac{d\theta}{(\theta_0{}^2 - \theta^2)^{1/2}} = \left(\frac{g}{L}\right)^{1/2} dt. \tag{11}$$

This form is convenient for integration.

If the initial condition or the phase of the motion is such that θ has the value θ_1 at $t = 0$, then

$$\int_{\theta_1}^{\theta} \frac{d\theta}{(\theta_0{}^2 - \theta^2)^{1/2}} = \left(\frac{g}{L}\right)^{1/2}\int_0^t dt. \tag{12}$$

The integral on the left is elementary [Dwight 320.01]:

$$\begin{aligned}\int_{\theta_1}^{\theta} \frac{d\theta}{(\theta_0{}^2 - \theta^2)^{1/2}} &= \left[\sin^{-1}\frac{\theta}{\theta_0}\right]_{\theta_1}^{\theta} = \sin^{-1}\frac{\theta}{\theta_0} - \sin^{-1}\frac{\theta_1}{\theta_0} \\ &= \left(\frac{g}{L}\right)^{1/2} t. \end{aligned} \tag{13}$$

We know that $\sin\,\sin^{-1}(\theta/\theta_0) = \theta/\theta_0$, so that we may rewrite (13) as

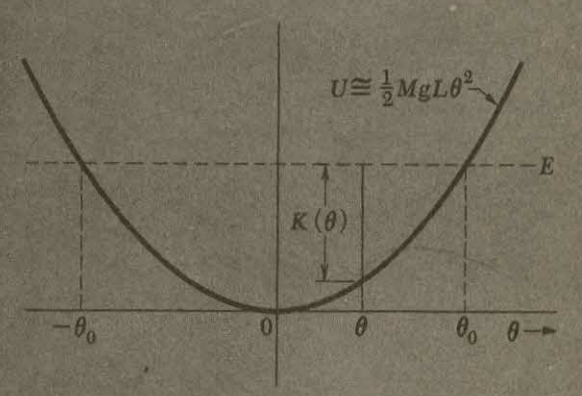
$$\frac{\theta}{\theta_0} = \sin\left[\left(\frac{g}{L}\right)^{1/2} t + \sin^{-1}\frac{\theta_1}{\theta_0}\right], \tag{14}$$

or

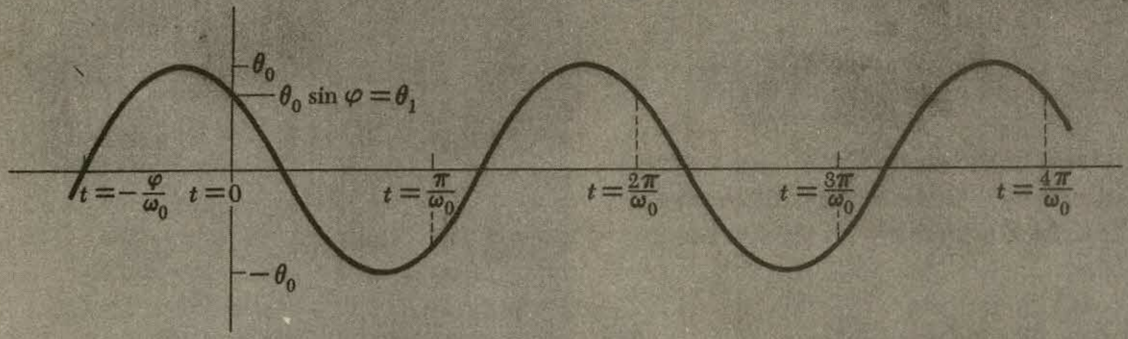
$$\theta = \theta_0 \sin\,(\omega_0 t + \varphi), \tag{15}$$

where we identify the angular frequency ω_0 and phase φ with

$$\omega_0 = \left(\frac{g}{L}\right)^{1/2}; \qquad \varphi = \sin^{-1}\frac{\theta_1}{\theta_0}. \tag{16}$$



The pendulum oscillates between the limits θ_0 and $-\theta_0$. At these "turning points," $K = 0$ and $U = E$. At $\theta = 0$, $U = 0$ and $K = E$. For $\theta \ll 1$ radian, $U \cong \frac{1}{2}MgL\,\theta^2$.



The function $\theta = \theta_0 \sin(\omega_0 t + \varphi)$ plotted vs t. After one full period of oscillation $2\pi/\omega_0$ has elapsed, the function repeats itself. The value of θ at $t = 0$ is denoted by θ_1 and is equal to $\theta_0 \sin \varphi$. That is, φ is chosen to give the desired value of θ_1. We call θ_0 the *amplitude* of the motion and we call φ the *phase* of the motion.

Here φ is a constant of the motion. Even though it has the dimensions of an angle, it is not an angle you can visualize immediately. The reader must get the three angular quantities θ_0, θ_1, and φ straight in his mind:

1. θ_0 is the maximum amplitude of oscillation.
2. θ_1 is the angle at which the motion started at $t = 0$. The pendulum may be set into motion from $\theta = \theta_1$ either by letting it go freely or by giving it some initial angular velocity. How we do it will affect θ_0.
3. In general, φ is no particular angle in the motion of the oscillator to which we can point. It is derived from θ_1 and θ_2, and it simply makes the proper correction to $\theta = \theta_0 \sin \omega_0 t$ if the motion is not started from $\theta_1 = 0$. The specific initial conditions tell us the value of φ through (16).

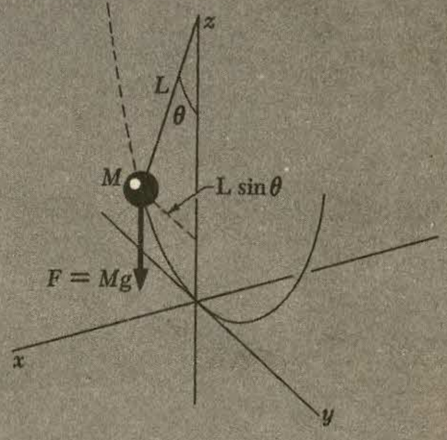
The symbol ω_0 is often used to denote the angular frequency of the natural or free motion of an oscillating system. The subscript "naught" on the ω has nothing to do with $t = 0$.

The angular frequency† ω_0 is related to the frequency f_0 of the free oscillation of the pendulum.

$$f_0 = \frac{\omega_0}{2\pi} = \frac{(g/L)^{1/2}}{2\pi}. \tag{17}$$

If $L = 100$ cm, we have $\omega_0 \approx (980/100)^{1/2} \approx 3$ rad/sec. The frequency is independent of the mass M and the amplitude θ_0 of the motion, provided that $\theta_0 \ll 1$. Note that there is no way the mass could enter into the right-hand side of (17) and still give a quantity with the dimensions of frequency.

We can also solve the problem of the simple pendulum by an approach starting with the equation of motion. In setting up (10) we used the law of conservation of energy in the form (5). Notice that (10) is a first-order differential equation, and we had to carry out only one integration with respect to time to obtain the result (14). The equation of motion is a second-order differential equation, as we see below. To solve for the deflection angle we have to integrate twice with respect to time. It is good to remember that the explicit use of energy conservation can often save mathematical labor by eliminating one integration.



The pendulum oscillates in the yz plane. The force due to gravity on M is $F = Mg$, in the $-z$ direction. The torque N_x arising from this force is $MgL \sin\theta$, in the $+x$ direction.

Let us take the x axis normal to the plane of the motion. The torque N_x due to gravity is, with $F = Mg$,

$$N_x = (\mathbf{r} \times \mathbf{F})_x = LMg \sin\theta, \tag{18}$$

taken about the pivot of the pendulum. The angular momentum J_x about the same point is, with the linear momentum $p = ML\dot{\theta}$,

$$J_x = (\mathbf{r} \times \mathbf{p})_x = -ML^2\dot{\theta}. \tag{19}$$

We know from Chap. 6 that the rate of change of angular momentum is equal to the torque:

$$ML^2\ddot{\theta} = -LMg \sin\theta, \tag{20}$$

so that the equation of motion of the pendulum is

$$\ddot{\theta} + \frac{g}{L}\sin\theta = 0. \tag{21}$$

---

† We shall often refer to the angular frequency simply as the frequency. Many physicists do this, and no particular confusion is caused. The use of the symbol ω rather than f or ν will usually identify a quantity as an angular frequency. As for numerical values, ν and f are usually given in cycles per second (cps); ω is given in radians per second or simply as $\sec^{-1}$, with the radians understood. A radian is dimensionless. Also it is rather common to differentiate between them by expressing the frequency ν in vibrations/sec, cycles/sec, or revolutions/sec and the angular frequency ω in radians/sec. Both have the dimensions $\sec^{-1}$.

In the limit $\theta \ll 1$ of small deflection angles we approximate $\sin\theta$ by θ:

$$\boxed{\ddot{\theta} + \frac{g}{L}\theta = 0.} \tag{22}$$

This is the equation of motion of a *harmonic oscillator* of angular frequency

$$\omega_0 = \left(\frac{g}{L}\right)^{1/2}. \tag{23}$$

To prove this statement, simply note that our result (15), or any linear combination of $\sin\omega_0 t$ and $\cos\omega_0 t$ is a solution of (22). We try (15):

$$\theta = \theta_0 \sin(\omega_0 t + \varphi); \tag{24}$$

$$\dot{\theta} = \omega_0 \theta_0 \cos(\omega_0 t + \varphi); \tag{25}$$

$$\ddot{\theta} = -\omega_0^2 \theta_0 \sin(\omega_0 t + \varphi). \tag{26}$$

With (24) and (26), we see that (22) reduces to

$$-\omega_0^2 \theta_0 \sin(\omega_0 t + \varphi) + \frac{g}{L}\theta_0 \sin(\omega_0 t + \varphi) = 0, \tag{27}$$

which agrees with (23) if we take $\omega_0^2 = g/L$.

The arithmetic is slightly neater if we take

$$\theta = \theta_0 e^{i(\omega_0 t + \varphi)}; \tag{28}$$

then

$$\dot{\theta} = i\omega_0\theta_0 e^{i(\omega_0 t + \varphi)} = i\omega_0\theta; \tag{29}$$

$$\ddot{\theta} = i\omega_0\dot{\theta} = (i\omega_0)^2\theta = -\omega_0^2\theta. \tag{30}$$

We see that (30) is exactly of the form (22), if $\omega_0^2 = g/L$.

We see that both (24) and (28) satisfy the equation of motion (22). Which is actually the correct solution? The answer is that (24) is the correct physical solution giving the angle which the arm of the pendulum makes at any time t. Equation (28) has an unphysical aspect as it stands because it contains the imaginary quantity i. When solving an equation of motion with complex quantities (which is sometimes easier mathematically), we must remember that in the end we take *either* the real or the imaginary part to obtain the physical solution. Note that the imaginary part of (28) is indeed (24), so that (28) *contains* the correct solution.

Example. Nonlinear effects. We consider now a pendulum which is oscillating with an amplitude so large that we may not neglect the θ^3 term in the expansion of sin θ, as we did above in (22). What is the effect on the motion of the pendulum of the term in θ^3? This is an elementary example of an anharmonic oscillator. Anharmonic or nonlinear problems are usually difficult to solve exactly (except by computers), but approximate solutions are often adequate to give us a good idea of what is happening.

The expansion of sin θ to terms of order θ^3, commonly expressed as "expansion to $O(\theta^3)$," is

$$\sin\theta = \theta - \tfrac{1}{6}\theta^3 + \cdots, \tag{31}$$

so that the equation of motion (21) becomes, to this order,

$$\frac{d^2\theta}{dt^2} + \omega_0{}^2\theta - \frac{\omega_0{}^2}{6}\theta^3 = 0, \tag{32}$$

where $\omega_0{}^2$ denotes the quantity g/L. This is the equation of motion of an *anharmonic oscillator.*

We see if we can find an approximate solution to (32) of the form

$$\theta = \theta_0 \sin \omega t + \epsilon\theta_0 \sin 3\omega t, \tag{33}$$

where ϵ is a dimensionless constant expected to be much less than 1 for $\theta_0 \ll 1$.

That is, we shall see if the motion can be represented approximately (or exactly—we don't know yet!) as the superposition of two different motions, one in sin ωt and the other in sin $3\omega t$. The presence of a term in sin $3\omega t$ is suggested by the trigonometric identity [Dwight 403.03]:

$$\sin^3 x \equiv \tfrac{3}{4}\sin x - \tfrac{1}{4}\sin 3x. \tag{34}$$

Thus the θ^3 term in the differential equation (32) will generate from the cube of sin ωt a term in sin $3\omega t$. To satisfy the differential equation we are forced to add to sin ωt a term such as ϵ sin $3\omega t$ just to cancel the sin $3\omega t$ term generated by θ^3. Going further, the new ϵ sin $3\omega t$ term in the trial solution will generate, on being cubed, a term in ϵ^3 sin $9\omega t$, and so on. There is no apparent reason why the process should stop, but if $\epsilon \ll 1$ it may be expected to converge rapidly because higher and higher powers of ϵ are involved as factors in the higher-frequency terms.

It is now evident that (33) can only be an approximate solution at best. It remains for us to determine ϵ, and also ω; while ω must reduce to ω_0 at small amplitudes, it may differ at large amplitudes. For simplicity we suppose that $\theta = 0$ at $t = 0$.

An approximate solution of this type to a differential equation is called a perturbation solution, because one term in the differential equation perturbs the motion which would occur without that term. As you have seen, we arrived at the form of (33) by guided guesswork. It is easy enough to try several guesses and to reject the ones which do not work.

We have from (33)

$$\begin{aligned} \ddot{\theta} &= -\omega^2\theta_0 \sin \omega t - 9\omega^2\epsilon\theta_0 \sin 3\omega t; \\ \theta^3 &= \theta_0{}^3(\sin^3 \omega t + 3\epsilon \sin^2 \omega t \sin 3\omega t + \cdots), \end{aligned} \tag{35}$$

where we have discarded the terms of order ϵ^2 and ϵ^3 because of our assumption that we can find a solution with $\epsilon \ll 1$. Then the terms of (32) become, using the trigonometric identity (34),

$$\begin{aligned} \ddot{\theta} &= -\omega^2\,\theta_0 \sin \omega t - 9\omega^2\,\epsilon\theta_0 \sin 3\omega t; \\ \omega_0{}^2\,\theta &= +\omega_0{}^2\,\theta_0 \sin \omega t + \omega_0{}^2\,\epsilon\theta_0 \sin 3\omega t; \\ -\frac{1}{6}\omega_0{}^2\,\theta^3 &= -\frac{3\omega_0{}^2}{24}\theta_0{}^3 \sin \omega t + \frac{\omega_0{}^2}{24}\theta_0{}^3 \sin 3\omega t \\ &\qquad - \frac{\omega_0{}^2}{2}\theta_0{}^3\,\epsilon \sin^2 \omega t \sin 3\omega t. \end{aligned} \tag{36}$$

Now add vertically the terms in (36). The sum on the left-hand side is equal to zero according to (32). If (33) is to be an approximate solution of (32) for all time t, it is necessary that the coefficients of $\sin \omega t$ and $\sin 3\omega t$ vanish separately on the right-hand side of (36). For suppose the coefficients did not vanish; then we would have an expression of the form $A \sin \omega t + B \sin 3\omega t = 0$, where A and B are constants. But such an equation *cannot* be satisfied at all times t, hence A and B must each be zero. By cutting off our assumed solution (33) at $3\omega t$ we have not included all the terms or frequencies which may occur, but we have included the most important ones.

The requirement that the coefficients of $\sin \omega t$ in (36) should sum to zero is that

$$-\omega^2 + \omega_0{}^2 - \tfrac{3}{24}\omega_0{}^2\theta_0{}^2 = 0, \tag{37}$$

or

$$\omega^2 = \omega_0{}^2(1 - \tfrac{1}{8}\theta_0{}^2); \qquad \omega \cong \omega_0(1 - \tfrac{1}{16}\theta_0{}^2), \tag{38}$$

using the binomial expansion for the square root. Equation (38) gives the dependence of ω on θ_0. Here ω_0 is the limit of ω as $\theta_0 \to 0$; that is, the small amplitude limit. For $\theta_0 = 0.3$ rad, the fractional frequency shift is $\Delta\omega/\omega \simeq -10^{-2}$, where $\Delta\omega \equiv \omega - \omega_0$. Note that the frequency of the pendulum at large amplitudes does depend on the amplitude.

The solution (33) also contains a term in sin $3\omega t$. The amplitude of this term relative to the amplitude of the term in sin ωt is ϵ, which is determined by the condition that the coefficient of the term in sin $3\omega t$ in (36) vanish:

$$-9\omega^2\epsilon + \omega_0{}^2\epsilon + \frac{\omega_0{}^2}{24}\theta_0{}^2 = 0. \tag{39}$$

If we set $\omega^2 \cong \omega_0{}^2$, then (39) reduces to

$$\epsilon \cong \frac{\theta_0{}^2}{192}. \tag{40}$$

We think of ϵ as giving the fractional admixture of the sin $3\omega t$ term in a solution for θ dominated by the sin ωt term. For $\theta_0 = 0.3$ rad, we have $\epsilon \simeq 10^{-3}$, which is very small. The coefficient of the term $\sin^2 \omega t \sin 3\omega t$ in (36) is small by $O(\epsilon)$ or by $O(\theta_0{}^2)$, compared with the terms we have used. We have neglected this term in our approximation.

Why did we not include in (33) a term in sin $2\omega t$? Try for yourself a solution of the form

$$\theta = \theta_0 \sin \omega t + \eta\theta_0 \sin 2\omega t, \tag{41}$$

and see what happens. You will find $\eta = 0$. The pendulum generates chiefly third harmonics, i.e., terms in sin $3\omega t$, and not second harmonics. The situation would be different for a device for which the equation of motion included a term in θ^2.

What is the frequency of the pendulum at large amplitudes? There is no single frequency in the motion. We have seen that the most important term (the largest component) is in sin ωt, and we say that ω is the *fundamental frequency* of the pendulum. To our approximation ω is given by (38). The term in sin $3\omega t$ is called the *third harmonic* of the fundamental frequency. Our argument following (33) suggests that an infinite number of harmonics are present in the exact motion, but that most of these are very small. The amplitude in (33) of the fundamental component of the motion is θ_0; the amplitude of the third harmonic component is $\epsilon\theta_0$.

Mass on a Spring. We considered in Chap. 5 a spring which satisfied Hooke's law of force:

$$F_x = -Cx, \tag{42}$$

where x is the coordinate of the end of the spring. If a mass M is attached to the end of the spring and the spring itself treated as massless, the equation of motion of the system is

$$M\ddot{x} = -Cx; \qquad \ddot{x} + \frac{C}{M}x = 0. \tag{43}$$

The solution of (43) is like that of (22). This equation is readily seen to have solutions of the form†

$$x = A \sin(\omega_0 t + \varphi), \tag{44}$$

where A and φ are the amplitude and phase constants. Now we show that (44) is a solution of (43). We form

$$\dot{x} = \omega_0 A \cos(\omega_0 t + \varphi); \qquad \ddot{x} = -\omega_0^2 A \sin(\omega_0 t + \varphi), \tag{45}$$

so that from (44) and (45) we see that

$$\ddot{x} + \omega_0^2 x = 0. \tag{46}$$

This is identical with (43) if the angular frequency ω_0 is given by

$$\omega_0 = \left(\frac{C}{M}\right)^{1/2} \tag{47}$$

The motion is harmonic with angular frequency ω_0. The amplitude is A. The phase φ is determined by the values of x and $\dot{x}$ at $t = 0$. Equations (44) and (45) yield at $t = 0$ the values

$$x_0 = A \sin \varphi \qquad \text{and} \qquad v_0 = \omega_0 A \cos \varphi;$$

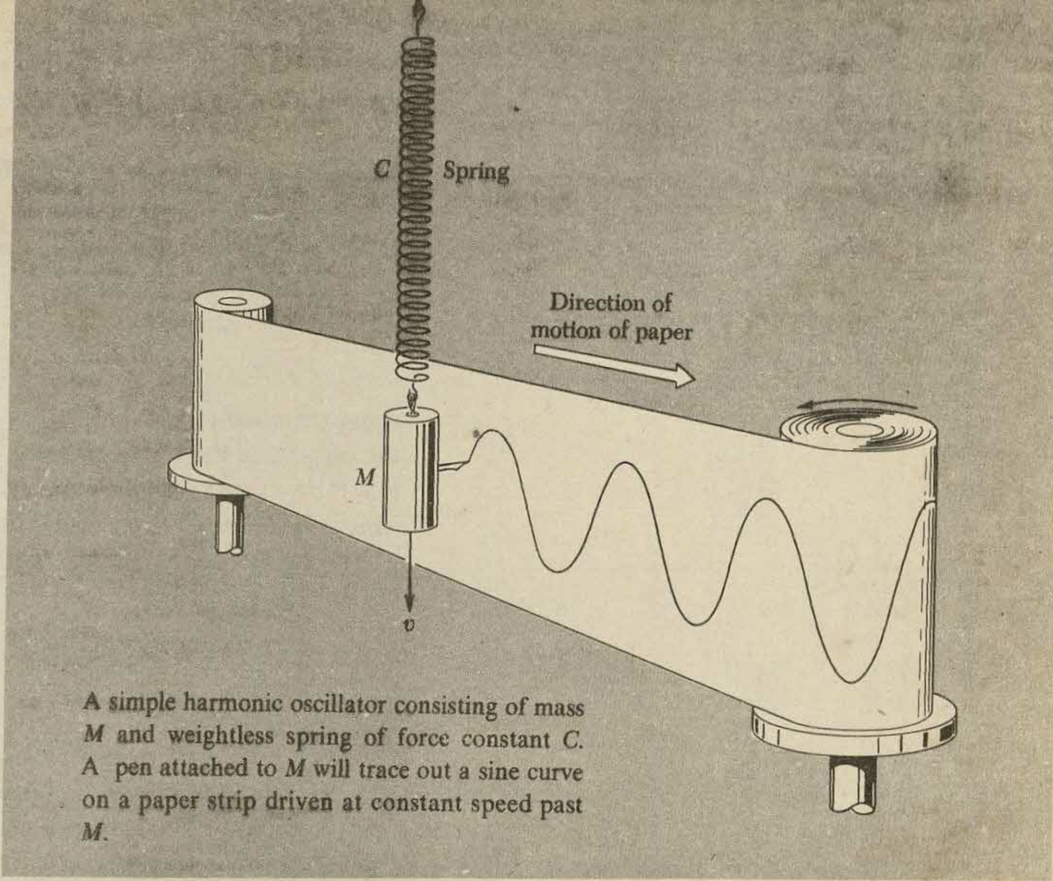
these two relations may be solved for A and φ. If $\varphi = \frac{1}{2}\pi$, then

$$x = A \sin(\omega_0 t + \tfrac{1}{2}\pi) = A \cos \omega_0 t, \tag{48}$$

because the sine and cosine differ only by a phase shift of $\frac{1}{2}\pi$. We see from (44) that $\varphi = \frac{1}{2}\pi$ if at $t = 0$ the spring is at its maximum displacement.

---

† We could equally well choose solutions of the form $x = A \cos(\omega_0 t + \varphi)$ or $x = B \cos \omega_0 t + D \sin \omega_0 t$. The selection of form is usually a matter of convenience.



A simple harmonic oscillator consisting of mass M and weightless spring of force constant C. A pen attached to M will trace out a sine curve on a paper strip driven at constant speed past M.

Example. Average kinetic and potential energy. Calculate the average over time of the kinetic and potential energy of a harmonic oscillator.

The kinetic energy is

$$K = \tfrac{1}{2}M\dot{x}^2 = \tfrac{1}{2}M[\omega_0 A \cos(\omega_0 t + \varphi)]^2, \tag{49}$$

using the solution (44). The time average† of the kinetic energy over one period T of the motion is

$$\langle K \rangle = \frac{\int_0^T K\,dt}{T} = \frac{1}{2}M\omega_0{}^2A^2\,\frac{\int_0^{2\pi/\omega_0}\cos^2(\omega_0 t + \varphi)\,dt}{2\pi/\omega_0}, \tag{50}$$

---

† We use angle brackets $\langle \ldots . \rangle$ to denote time average, as defined here by

$$\langle x \rangle = \lim_{t\to\infty} \frac{1}{t}\int_0^t x(t)\,dt.$$

Clearly, for a quantity $q(t)$ that repeats itself with period T, the time average can also be written

$$\langle q \rangle = \frac{1}{T}\int_0^T q(t)\,dt.$$

where $2\pi/\omega_0$ is the period. Because the integral is extended over a complete period, it does not matter what value the phase φ has, and we may conveniently set $\varphi = 0$. Then, if we write $y = \omega_0 t$, we have

$$\frac{\omega_0}{2\pi}\int_0^{2\pi/\omega_0}\cos^2 \omega_0 t\,dt = \frac{1}{2\pi}\int_0^{2\pi}\cos^2 y\,dy = \frac{1}{2}, \tag{51}$$

using the identity $\cos^2 y = \frac{1}{2}(1 + \cos 2y)$ and observing that the integral of the term in $\cos 2y$ vanishes. We see from (51) that the average value of the cosine squared is equal to one-half. This is a result worth remembering. The same is true of the average of the sine squared.[†] From (50) and (51) we have for the average kinetic energy

$$\langle K\rangle = \tfrac{1}{4}M\omega_0{}^2A^2. \tag{52}$$

The potential energy is, with $x = A \sin \omega_0 t$,

$$U = \tfrac{1}{2}Cx^2 = \tfrac{1}{2}CA^2 \sin^2 \omega_0 t. \tag{53}$$

By analogy with (51) the average value of the sine squared is

$$\frac{\omega_0}{2\pi}\int_0^{2\pi/\omega_0}\sin^2 \omega_0 t\,dt = \frac{1}{2}, \tag{54}$$

so that, using $\omega_0{}^2 = C/M$ and the result (53),

$$\langle U\rangle = \tfrac{1}{4}CA^2 = \tfrac{1}{4}M\omega_0{}^2A^2. \tag{55}$$

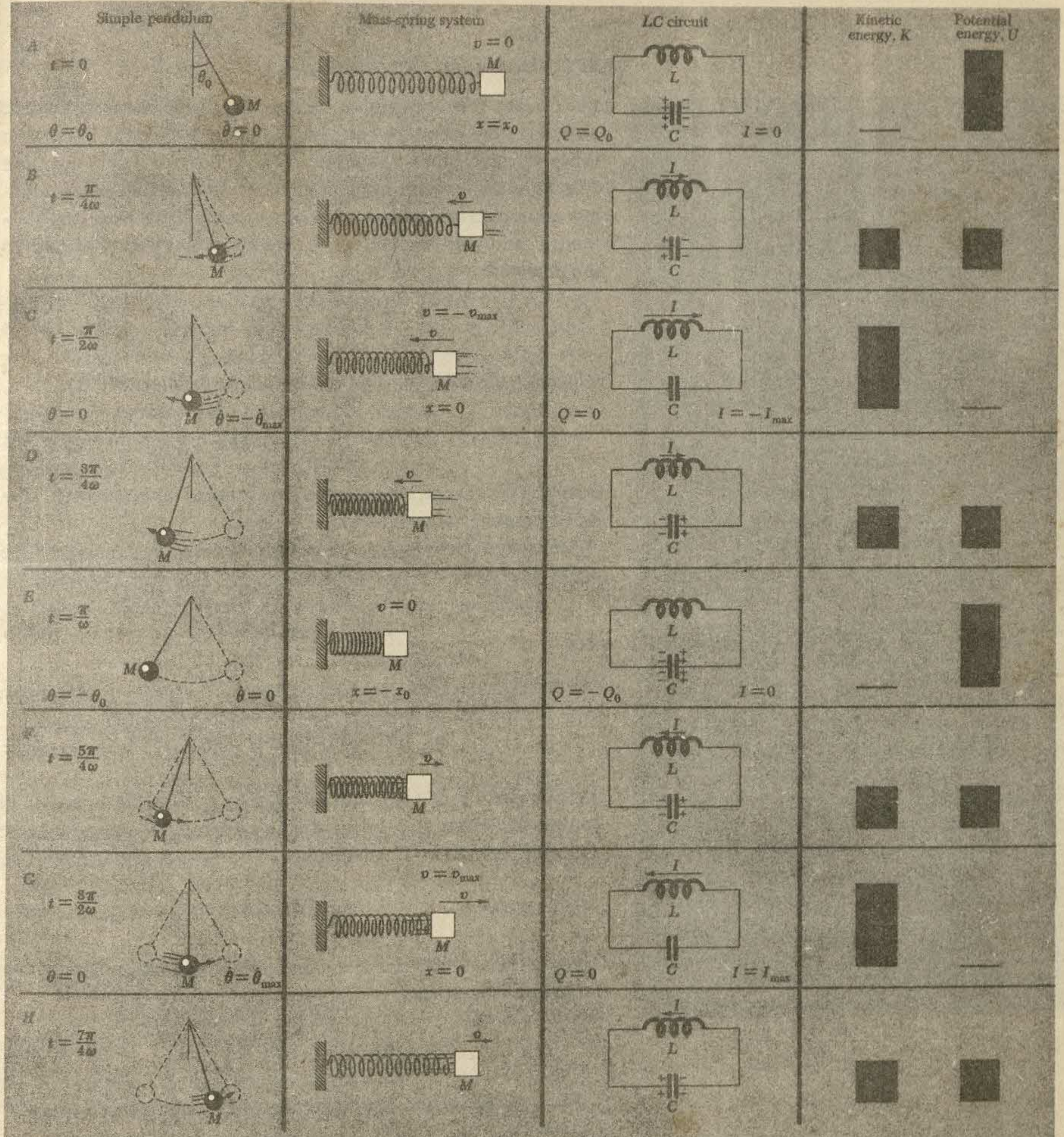
Thus $\langle U\rangle = \langle K\rangle$ and the total energy of the harmonic oscillator is

$$\boxed{E = \langle K\rangle + \langle U\rangle = \tfrac{1}{2}M\omega_0{}^2A^2.} \tag{56}$$

Note that $E = \langle E\rangle$ because the total energy is a constant of the motion.

The equality of the average kinetic energy and potential energies is a special property of the harmonic oscillator. This property does not hold in general for anharmonic oscillators. It will be shown later to hold for weakly damped oscillators.

[†] We can derive these results easily from the trigonometric identity $\sin^2\varphi + \cos^2\varphi = 1$. On averaging over 2π radians, we have $\langle\sin^2\varphi\rangle + \langle\cos^2\varphi\rangle = 1$. But the only difference between $\sin\varphi$ and $\cos\varphi$ is a matter of a $\pi/2$ phase shift, and thus $\langle\sin^2\varphi\rangle = \langle\cos^2\varphi\rangle = \frac{1}{2}$. This type of argument can also be applied to the angular average value of $\langle x^2\rangle$ over the surface of a sphere. If $x^2 + y^2 + z^2 = r^2$, then $\langle x^2\rangle + \langle y^2\rangle + \langle z^2\rangle = r^2$. Because the sphere is symmetric with respect to x, y, z we must have $\langle x^2\rangle = \langle y^2\rangle = \langle z^2\rangle = \frac{1}{3}r^2$. This result may be confirmed by direct calculation.



Three separate harmonic oscillators with the same period: a simple pendulum, a mass-spring system, and an LC circuit. Time increases from Fig. A to Fig. H; the next cycle begins again with Fig. A.

where $2\pi/\omega_0$ is the period. Because the integral is extended over a complete period, it does not matter what value the phase φ has, and we may conveniently set $\varphi = 0$. Then, if we write $y = \omega_0 t$, we have

$$\frac{\omega_0}{2\pi}\int_0^{2\pi/\omega_0} \cos^2 \omega_0 t \, dt = \frac{1}{2\pi}\int_0^{2\pi} \cos^2 y \, dy = \frac{1}{2}, \tag{51}$$

using the identity $\cos^2 y = \frac{1}{2}(1 + \cos 2y)$ and observing that the integral of the term in $\cos 2y$ vanishes. We see from (51) that the average value of the cosine squared is equal to one-half. This is a result worth remembering. The same is true of the average of the sine squared.† From (50) and (51) we have for the average kinetic energy

$$\langle K \rangle = \tfrac{1}{4}M\omega_0{}^2A^2. \tag{52}$$

The potential energy is, with $x = A \sin \omega_0 t$,

$$U = \tfrac{1}{2}Cx^2 = \tfrac{1}{2}CA^2 \sin^2 \omega_0 t. \tag{53}$$

By analogy with (51) the average value of the sine squared is

$$\frac{\omega_0}{2\pi}\int_0^{2\pi/\omega_0} \sin^2 \omega_0 t \, dt = \frac{1}{2}, \tag{54}$$

so that, using $\omega_0{}^2 = C/M$ and the result (53),

$$\langle U \rangle = \tfrac{1}{4}CA^2 = \tfrac{1}{4}M\omega_0{}^2A^2. \tag{55}$$

Thus $\langle U \rangle = \langle K \rangle$ and the total energy of the harmonic oscillator is

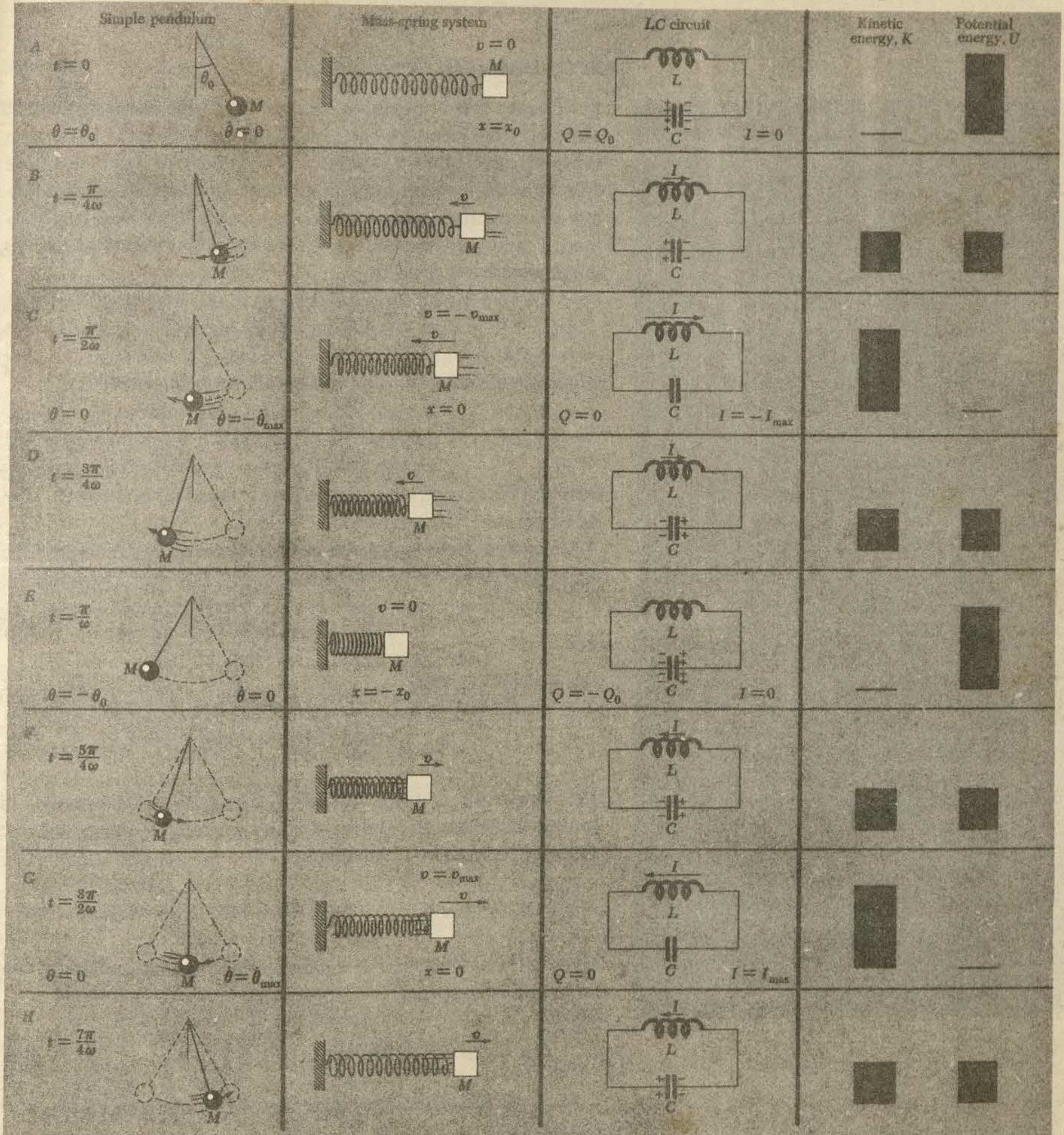
$$\boxed{E = \langle K \rangle + \langle U \rangle = \tfrac{1}{2}M\omega_0{}^2A^2.} \tag{56}$$

Note that $E = \langle E \rangle$ because the total energy is a constant of the motion.

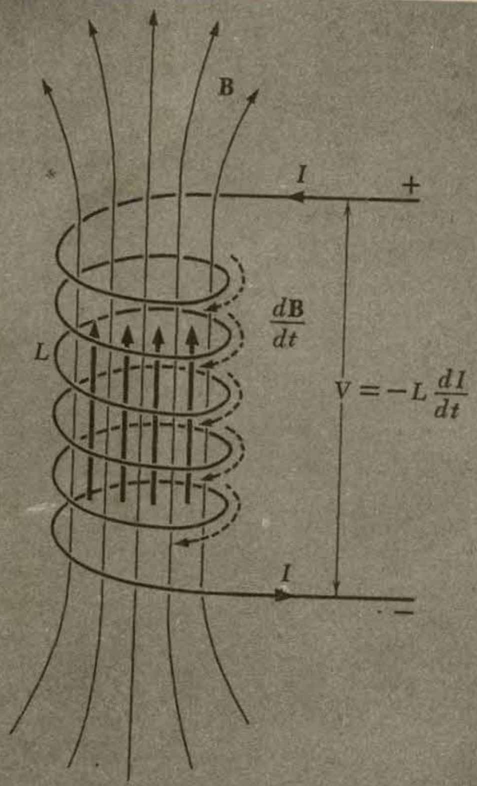
The equality of the average kinetic energy and potential energies is a special property of the harmonic oscillator. This property does not hold in general for anharmonic oscillators. It will be shown later to hold for weakly damped oscillators.

---

† We can derive these results easily from the trigonometric identity $\sin^2 \varphi + \cos^2 \varphi = 1$. On averaging over 2π radians, we have $\langle \sin^2 \varphi \rangle + \langle \cos^2 \varphi \rangle = 1$. But the only difference between $\sin \varphi$ and $\cos \varphi$ is a matter of a $\pi/2$ phase shift, and thus $\langle \sin^2 \varphi \rangle = \langle \cos^2 \varphi \rangle = \frac{1}{2}$. This type of argument can also be applied to the angular average value of $\langle x^2 \rangle$ over the surface of a sphere. If $x^2 + y^2 + z^2 = r^2$, then $\langle x^2 \rangle + \langle y^2 \rangle + \langle z^2 \rangle = r^2$. Because the sphere is symmetric with respect to x, y, z we must have $\langle x^2 \rangle = \langle y^2 \rangle = \langle z^2 \rangle = \frac{1}{3}r^2$. This result may be confirmed by direct calculation.



Three separate harmonic oscillators with the same period: a simple pendulum, a mass-spring system, and an LC circuit. Time increases from Fig. A to Fig. H; the next cycle begins again with Fig. A.



Inductor of inductance L carries current I. If I increases, $\mathbf{B}$ increases. Direction of $d\mathbf{B}/dt$ is shown by heavy arrows. According to Faraday's law an electric field is induced in the coil by the changing magnetic field. The direction of the electric field is shown by the dotted arrows. The total voltage across the coil is $V = \int \mathbf{E} \cdot d\mathbf{l}$. Since V increases in the opposite direction to dI/dt, we have $V = -L(dI/dt)$, where L is the constant of proportionality.

***LC* Circuit.** We anticipate here some of the electrical circuit analysis of the laboratory and of Vol. II. (This section along with Probs. 8 and 9 may be omitted by classes which do not have a background in electric circuit elements.) Consider a resistanceless series electrical circuit consisting of an inductance L, a capacitance C, and a driving voltage V. We know that the voltage across the inductance is given by

$$V_L = -L\frac{dI}{dt}, \tag{57}$$

where I is the current. The voltage across the capacitance C is

$$V_C = \frac{Q}{C} = -\frac{1}{C}\int I\,dt; \tag{58}$$

here Q is the charge, which is the time integral of the current. The sign is a matter of convention.

The sum of the voltages on passing around the circuit must be zero. Therefore $V + V_L + V_C = 0$, or

$$L\frac{dI}{dt} + \frac{1}{C}\int I\,dt = V, \tag{59}$$

or, using $Q = -\int I\,dt$,

$$L\frac{d^2Q}{dt^2} + \frac{1}{C}Q = -V. \tag{60}$$

On comparison of (59) with $M\ddot{x} + C_{\text{spring}}x = F$, the equation of motion of a mass on a spring driven by a force F, we see that the following quantities are analogous:

$$Q \leftrightarrow x; \qquad -V \leftrightarrow F; \qquad L \leftrightarrow M; \qquad \frac{1}{C} \leftrightarrow C_{\text{spring}}. \tag{61}$$

The solutions of (60) for $V = 0$ are of the form, by analogy to the solution for the spring,

$$I = I_0 \sin \omega_0 t; \qquad \omega_0 = \left(\frac{1}{LC}\right)^{1/2}. \tag{62}$$

The voltage across the inductance is found from (57) and (62) as

$$V_L = -L\dot{I} = -L\omega_0 \cos \omega_0 t. \tag{63}$$

The voltage across the capacitance is

$$V_C = -\frac{1}{C}\int I\,dt = \frac{1}{\omega_0 C}\cos\omega_0 t. \tag{64}$$

From the definition (62) of ω_0 we see that the amplitudes of V_L and V_C are equal.

Friction. Thus far we have neglected frictional effects on the harmonic oscillator. Friction damps out the free motion of the oscillator. The solutions are more realistic when friction is included in the equation of motion. How do we introduce friction into the equation of motion $M\ddot{x} = 0$ of a free particle? Friction is a retarding force on the particle. If it is the only force acting on the particle, then by Newton's second law

$$M\ddot{x} = F_f. \tag{65}$$

The frictional force must oppose the velocity of the particle and in its simplest form is linear proportional to the velocity. (This is not obvious; examples are discussed below). Let us consider

$$F_f = -\gamma\dot{x}, \tag{66}$$

where γ is a positive constant, called the damping coefficient. Thus the equation of motion of a particle moving under the action of frictional forces only is

$$M\ddot{x} + \gamma\dot{x} = 0. \tag{67}$$

It is sometimes useful to define a constant τ, called the *relaxation time,* by the relation

$$\gamma \equiv \frac{M}{\tau}, \tag{68}$$

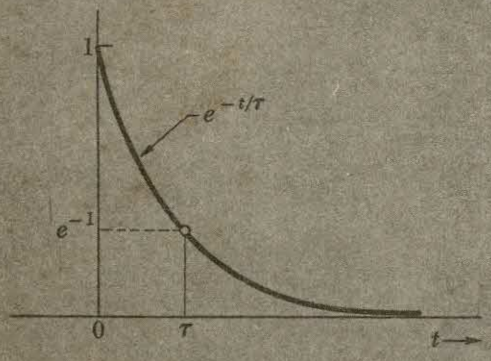
for then (67) becomes

$$M\left(\frac{d^2x}{dt^2} + \frac{1}{\tau}\frac{dx}{dt}\right) = 0. \tag{69}$$

We see that τ has the dimensions of time.

In terms of the velocity $v \equiv \dot{x}$, this equation becomes

$$\dot{v} + \frac{1}{\tau}v = 0. \tag{70}$$



The function $e^{-t/\tau}$ is plotted vs t.

This is a very important differential equation. It may be rewritten as

$$\frac{dv}{v} = -\frac{dt}{\tau}, \qquad \text{or} \qquad \int_{v_0}^{v} \frac{dv}{v} = -\frac{1}{\tau}\int_0^t dt, \tag{71}$$

where v_0 is the velocity at $t = 0$. On integrating,

$$\log v - \log v_0 = -\frac{t}{\tau}, \qquad \text{or} \qquad \log \frac{v}{v_0} = -\frac{t}{\tau}. \tag{72}$$

On raising both sides to powers of e, we have

$$v(t) = v_0 e^{-t/\tau}. \tag{73}$$

The velocity decreases exponentially with time; we say that the velocity is damped with a time constant τ.

The decay of the kinetic energy K of a free particle is given from (73) by

$$K = \tfrac{1}{2}Mv^2 = \tfrac{1}{2}Mv_0{}^2 e^{-2t/\tau} = K_0 e^{-2t/\tau}. \tag{74}$$

We see on differentiation of (74) that

$$\dot{K} = -\frac{2}{\tau}K. \tag{75}$$

The effective relaxation time for the kinetic energy is one-half that for the velocity.

What sort of mechanism leads to a damping force of the form (66)? This form is an idealization which can be realized under certain conditions. An ohmic electrical resistance furnishes a damping or frictional element of exactly the form (66). The voltage drop V_R across an ideal resistor is related to the current by Ohm's law:

$$V_R = IR. \tag{76}$$

The equation of motion of an undriven LR circuit is given by (57) and (76):

$$L\dot{I} + RI = 0, \tag{77}$$

which is exactly of the form (70) with $\tau = L/R$. A discussion of the microscopic mechanism leading to Ohm's law is given in Vol. II.

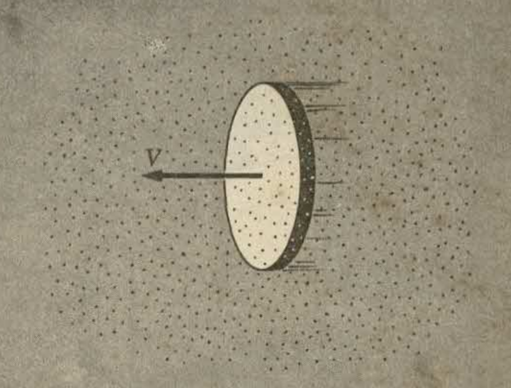
A good representation of the damping force $F_f = -\gamma\dot{x}$ is also provided by a flat plate moving normal to its plane through a gas at very low pressure, provided the speed V of the plate is much slower† than the average speed v of the molecules of the gas. The pressure must be sufficiently low that we can neglect the collisions of the molecules with each other. The rate at which molecules strike the plate is proportional to the relative velocity of the incoming molecules and the plate. Suppose that the molecules move only in one dimension. On one side of the plate the relative velocity is $v + V$; on the other side it is $v - V$. The pressure is proportional to the rate at which molecules strike the plate times the average momentum transfer per molecule. The average momentum transfer is itself proportional to the relative velocity, so that the pressures P_1, P_2 on the two sides of the plate are

$$P_1 \propto (v + V)^2; \qquad P_2 \propto (v - V)^2. \tag{78}$$

The net pressure P is the difference of the pressures on the opposite sides of the plate:

$$P = P_1 - P_2 \propto 4vV, \tag{79}$$

so that the drag (the net force on the moving plate) is directly proportional to the speed V of the plate. The direction of the drag can be seen to oppose the motion of the plate.



A flat plate moving normal to its plane through a gas at very low pressure is subjected to a retarding force proportional to its velocity V, if V is much less than the average speed of the molecules of the gas.

Damped Harmonic Oscillator. If we include the damping force, the total force acting on an undriven harmonic oscillator in one dimension is

$$F_{\text{spring}} + F_{\text{friction}} = -Cx - \gamma\dot{x}. \tag{80}$$

We set this equal to $M\ddot{x}$ and obtain the equation of motion

$$M\ddot{x} + \gamma\dot{x} + Cx = 0. \tag{81}$$

This is still a linear equation. We may rewrite (81) in the form

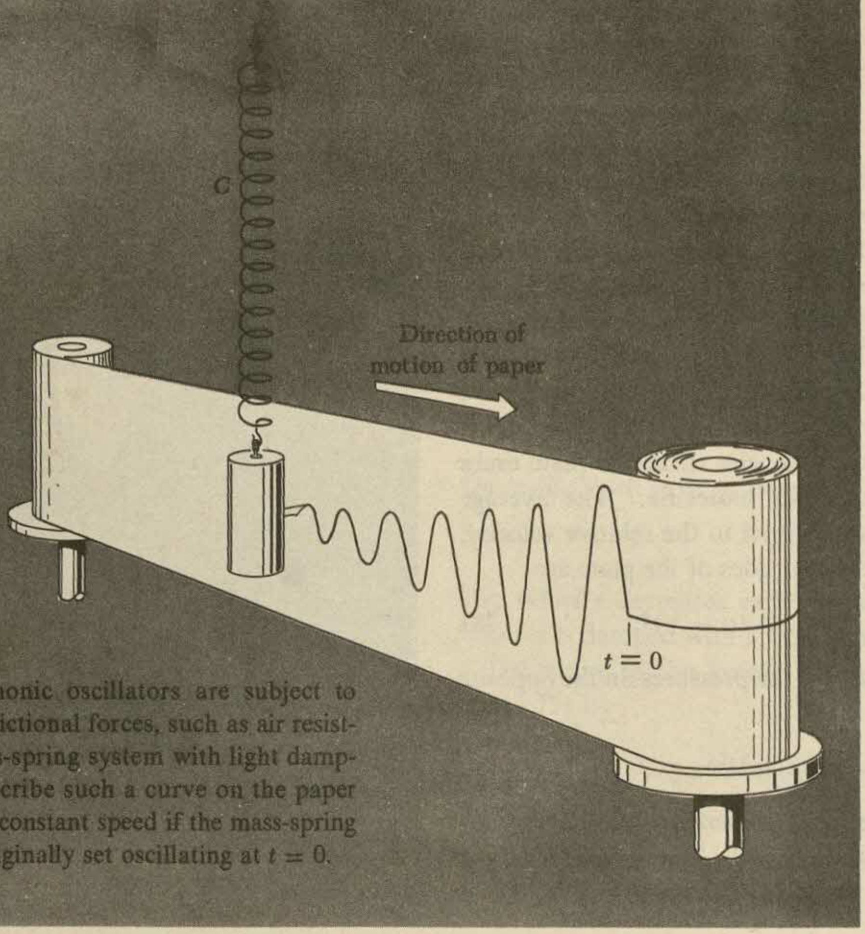
$$\boxed{\ddot{x} + \frac{1}{\tau}\dot{x} + \omega_0{}^2x = 0,} \tag{82}$$

where

$$\frac{1}{\tau} = \frac{\gamma}{M}; \qquad \omega_0{}^2 = \frac{C}{M}. \tag{83}$$

---

† The situation in which the motion of the plate is fast in comparison with molecular velocities was considered in the Advanced Topic on the slowing down of meteoroids, in Chap. 6. In both problems we assume for simplicity that the mean free path of the molecules is large in comparison with the dimensions of the plate.


C
Direction of motion of paper
$t = 0$

All real harmonic oscillators are subject to damping by frictional forces, such as air resistance. A mass-spring system with light damping would describe such a curve on the paper tape driven at constant speed if the mass-spring system was originally set oscillating at $t = 0$.

We look for solutions of (82) in the form of damped sinusoidal oscillations:

$$x = x_0 e^{-\beta t} \sin \omega t, \tag{84}$$

where β and ω are to be determined. This solution is suggested by combining (44) and (73). We could equally well have taken $x = x_0 e^{-\beta t} \cos \omega t$. We note on differentiating (84) that

$$\frac{dx}{dt} = -\beta x_0 e^{-\beta t} \sin \omega t + \omega x_0 e^{-\beta t} \cos \omega t; \tag{85}$$

$$\frac{d^2x}{dt^2} = \beta^2 x_0 e^{-\beta t} \sin \omega t - 2\omega\beta x_0 e^{-\beta t} \cos \omega t - \omega^2 x_0 e^{-\beta t} \sin \omega t. \tag{86}$$

The differential equation (82) becomes, on collecting terms,

$$\left(\beta^2 - \omega^2 + \omega_0{}^2 - \frac{\beta}{\tau}\right) x_0 e^{-\beta t} \sin \omega t + \left(-2\omega\beta + \frac{\omega}{\tau}\right) x_0 e^{-\beta t} \cos \omega t = 0. \tag{87}$$

Now observe that in (87) the coefficient of the term in cos ωt will be zero if

$$\beta = \frac{1}{2\tau}, \tag{88}$$

and the coefficient of the term in sin ωt will be zero if

$$\omega^2 = \omega_0{}^2 + \beta^2 - \frac{\beta}{\tau} = \omega_0{}^2 - \left(\frac{1}{2\tau}\right)^2; \tag{89}$$

or

$$\omega = \left[\omega_0{}^2 - \left(\frac{1}{2\tau}\right)^2\right]^{1/2} = \omega_0\left[1 - \left(\frac{1}{2\omega_0\tau}\right)^2\right]^{1/2}. \tag{90}$$

Friction lowers the frequency. The frequence ω is equal to ω_0 only if the relaxation time is infinite (no damping).

If β and ω are given by (88) and (90), then (84) is a solution of the equation of motion (81). That is, the solution is

$$x = x_0 e^{-t/2\tau} \sin\left\{\omega_0 t\left[1 - \left(\frac{1}{2\omega_0\tau}\right)^2\right]^{1/2}\right\}. \tag{91}$$

If $\omega_0\tau \gg 1$, we have the limit of *low damping* in which (91) may be approximated by

$$x \cong x_0 e^{-t/2\tau} \sin \omega_0 t, \tag{92}$$

where ω_0 is the natural frequency of the undamped oscillation.

Example. Power dissipation. Calculate the rate of energy dissipation by a damped harmonic oscillator, in the weak damping limit with $\omega_0\tau \gg 1$, so that $\omega \cong \omega_0$.

The kinetic energy is $K = \frac{1}{2}M\dot{x}^2$. From the approximate solution (92) we have

$$\frac{dx}{dt} = -\frac{1}{2\tau}x_0 e^{-t/2\tau} \sin \omega_0 t + \omega_0 x_0 e^{-t/2\tau} \cos \omega_0 t. \tag{93}$$

Now

$$\left(\frac{dx}{dt}\right)^2 = \left(\frac{1}{2\tau}\right)^2 x_0{}^2 e^{-t/\tau} \sin^2 \omega_0 t + \omega_0{}^2 x_0{}^2 e^{-t/\tau} \cos^2 \omega_0 t$$
$$- \left(\frac{\omega_0}{\tau}\right) x_0{}^2 e^{-t/\tau} \sin \omega_0 t \cos \omega_0 t. \tag{94}$$

The integrals that arise when we take the time average of (94) are given in Dwight, 861.13 to 861.15. But for $\omega_0\tau \gg 1$, it is a good approximation to take the factor $e^{-t/\tau}$ in (94) outside the angle brackets which denote time average. We can do this to reasonable accuracy if the amplitude of the oscillation $x_0e^{-t/2\tau}$ does not change much in one cycle of the motion. We are left with the averages:

$$\langle\cos^2\theta\rangle = \langle\sin^2\theta\rangle = \tfrac{1}{2}; \qquad \langle\cos\theta\sin\theta\rangle = 0. \tag{95}$$

The last average, which is new to us, is of considerable importance:

$$\langle\cos\theta\sin\theta\rangle = \langle\tfrac{1}{2}\sin 2\theta\rangle = 0, \tag{96}$$

because the average of a sine or cosine is zero. Then the kinetic energy averaged over the time of one cycle is

$$\begin{aligned}\langle K\rangle &\cong \frac{1}{2}M\left[\left(\frac{1}{2\tau}\right)^2\langle\sin^2\omega_0t\rangle + \omega_0{}^2\langle\cos^2\omega_0t\rangle\right.\\ &\qquad\qquad\left. - \frac{\omega_0}{\tau}\langle\cos\omega_0t\sin\omega_0t\rangle\right]x_0{}^2e^{-t/\tau}\\ &\cong \frac{1}{4}M\left[\left(\frac{1}{2\tau}\right)^2 + \omega_0{}^2\right]x_0{}^2e^{-t/\tau},\end{aligned} \tag{97}$$

but $(1/2\tau)^2$ is assumed to be negligible in comparison to $\omega_0{}^2$, so that the average kinetic energy is

$$\langle K\rangle \cong \tfrac{1}{4}M\omega_0{}^2x_0{}^2e^{-t/\tau}. \tag{98}$$

We see that the average kinetic energy decays exponentially. The average potential energy is

$$\langle U\rangle = \tfrac{1}{2}M\omega_0{}^2x_0{}^2\langle e^{-t/\tau}\sin^2\omega_0t\rangle \cong \tfrac{1}{4}M\omega_0{}^2x_0{}^2e^{-t/\tau}. \tag{99}$$

The average power dissipation P is given by the negative of the rate of change of energy:

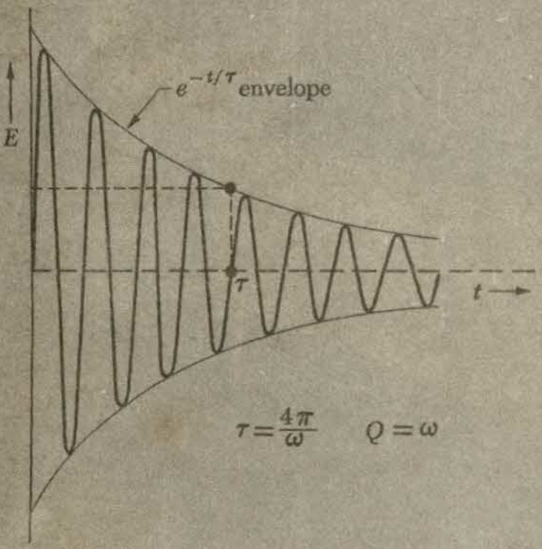
$$-\langle P\rangle = \frac{d}{dt}\langle E\rangle \cong \frac{d}{dt}\Big(\langle K\rangle + \langle U\rangle\Big) \cong -\frac{1}{\tau}\left(\frac{1}{2}M\omega_0{}^2x_0{}^2e^{-t/\tau}\right),$$

or

$$\langle P(t)\rangle = \frac{\langle E(t)\rangle}{\tau}. \tag{100}$$

We usually omit the $\langle\ \rangle$ on $P(t)$ when the meaning is clear.

The student may be surprised that the averages expressed in (98) and (99) contain the time t, as these are averages over time. We are viewing the motion of a damped oscillator over



Potential energy of oscillator with Q of 8π, plotted vs time. In time τ, in which four oscillations occur, the envelope of the oscillation decays to $1/e$ of its initial value.

many cycles, and what we have here is the average (kinetic or potential) energy *over one cycle at about the time t.* Because energy is being dissipated into heat, we expect the average energy (over a cycle) to decrease as more and more cycles are completed.

We expect to find that the power dissipation is equal to the negative of the average rate at which work is done by the frictional force $F_f = -\gamma\dot{x} = -(M/\tau)\dot{x}$. Using (93) and assuming that $\omega_0\tau \gg 1$ so that $e^{-t/\tau}$ can be taken out of the $\langle\ \rangle$ brackets, this average rate of doing work is

$$\begin{aligned}\langle F_f v\rangle &\cong -\frac{M}{\tau}\omega_0{}^2x_0{}^2e^{-t/\tau}\langle\cos^2\omega_0 t\rangle\\ &\cong -\frac{1}{2\tau}M\omega_0{}^2x_0{}^2e^{-t/\tau} \cong -\frac{E(t)}{\tau},\end{aligned} \tag{101}$$

in agreement with (100).

Quality Factor Q. The Q or *quality factor* of an oscillatory system is a term very widely used. The Q is defined as 2π times the ratio of the energy stored to the average energy loss per period:

$$Q = 2\pi\frac{\text{energy stored}}{\langle\text{energy loss in one period}\rangle} = \frac{2\pi E}{P/\nu} = \frac{E}{P/\omega}, \tag{102}$$

because the period is $1/\nu$ and $2\pi\nu = \omega$. The time of 1 radian of motion is $1/\omega$. Note that Q is dimensionless.

For the lightly damped harmonic oscillator ($\omega_0\tau \gg 1$) we have

$$Q \cong \frac{E}{E/\omega\tau} \cong \omega_0\tau \tag{103}$$

from (98), (99), and (100). We see that the value of $\omega_0\tau$ is indeed a good measure of the lack of damping of an oscillator. High $\omega_0\tau$ or high Q means that the oscillator is lightly damped. Note from (98) and (99) that the energy of an oscillator decays to e^{-1} of its initial value in the time τ; during this time the oscillator performs $\omega_0\tau/2\pi$ oscillations. Several representative values of Q are given in the accompanying table.

Several typical values of Q

(Wide variation may be expected)

Earth, for earthquake wave	250–1400
Copper cavity microwave resonator	10^4
Piano or violin string	10^3
Excited atom	10^7
Excited nucleus (Fe^{57})	3×10^{12}

Driven Harmonic Oscillator. We now consider in detail the driven or forced motion of a damped harmonic oscillator. *This is a problem of the greatest importance.* If besides friction there is an external force $F(t)$ applied to the oscillator, the equation of motion is

$$M\ddot{x} + \gamma\dot{x} + Cx = F(t), \tag{104}$$

or, in a more compact notation with $\tau \equiv M/\gamma$ and $\omega_0{}^2 \equiv C/M$,

$$\ddot{x} + \frac{1}{\tau}\dot{x} + \omega_0{}^2x = \frac{F(t)}{M}. \tag{105}$$

Here ω_0 is the *natural frequency* of the system in the absence of friction and in the absence of a driving force. When the system is driven at a different frequency $\omega(\neq\omega_0)$, we shall see that the response will be at the driving frequency and *not* at the natural frequency. But if the driving frequency is suddenly switched off, the system will revert to a damped oscillation at approximately the natural frequency, provided that the damping is low.

Suppose in (105) that

$$\frac{F(t)}{M} = \frac{F_0 \sin \omega t}{M} \equiv \alpha_0 \sin \omega t; \qquad \alpha_0 \equiv \frac{F_0}{M}, \tag{106}$$

so that the driving force is sinusoidal at frequency ω. This relation defines the quantity α_0. We introduce α_0 because it is more convenient to write F_0/M than F_0.

The steady-state (the state of the system after any transient effects have died down) response of the system will be precisely at the driving frequency. Otherwise the relative phase between force and response would change with time. This is an important feature of the result—the steady-state response of a driven harmonic oscillator (even with damping) is at the *driving frequency* and not at the natural frequency ω_0. No frequency other than the driving frequency will satisfy the equation of motion. By response we can mean either the displacement x or the velocity $\dot{x}$. We shall speak here of x as the response.

We look for a solution of (105) of the form

$$x = x_0 \sin(\omega t + \varphi), \tag{107}$$

where we have to solve the equation of motion for the values of the amplitude x_0 and the phase angle† φ. In (107) ω is the frequency of the driving force, *not* the natural frequency of the oscillator; and φ gives the phase between the driving force and the displacement of the oscillator. Thus φ here has quite a different meaning from what we encountered in the undriven undamped harmonic oscillator, where φ was related to the initial conditions. The initial conditions are irrelevant to the driven oscillator if only the steady state is considered.

---

† We need to allow the angle φ (called the phase angle of x relative to the force F) to be different from zero. No solution can be obtained if φ is left out. In speaking of a phase angle, be sure to say the phase of what, relative to what. In

It is worthwhile to define precisely what we mean by the phase φ between the displacement and driving force. Both the driving force and the displacement oscillate with simple harmonic motion. The cycle from maximum to maximum of both the force and displacement takes 360° or 2π radians. *The phase φ tells us by what angle the displacement reaches its maximum before the force.* For instance, suppose the force attains its greatest positive value at the instant when the displacement is zero and increasing in the positive direction. Then the displacement will lag the force by $\pi/2$ radians. But φ is defined as the angle by which x *leads* F, so that φ equals $-\pi/2$ in this instance.

Let us form the derivatives

$$\frac{dx}{dt} = \omega x_0 \cos(\omega t + \varphi); \qquad \frac{d^2x}{dt^2} = -\omega^2 x_0 \sin(\omega t + \varphi). \tag{108}$$

Then the equation of motion (105) is

$$(\omega_0{}^2 - \omega^2)\, x_0 \sin(\omega t + \varphi) + \frac{\omega}{\tau} x_0 \cos(\omega t + \varphi) = \alpha_0 \sin \omega t. \tag{109}$$

We simplify this by the trigonometric relations

$$\sin(\omega t + \varphi) = \sin \omega t \cos \varphi + \cos \omega t \sin \varphi; \tag{110}$$

$$\cos(\omega t + \varphi) = \cos \omega t \cos \varphi - \sin \omega t \sin \varphi. \tag{111}$$

Thus (109) becomes

$$\begin{aligned} &\left[(\omega_0{}^2 - \omega^2)\cos\varphi - \frac{\omega}{\tau}\sin\varphi\right] x_0 \sin \omega t \\ &\qquad + \left[(\omega_0{}^2 - \omega^2)\sin\varphi + \frac{\omega}{\tau}\cos\varphi\right] x_0 \cos \omega t = \alpha_0 \sin \omega t. \end{aligned} \tag{112}$$

---

electrical problems it is customary to speak of the phase of the current referred to the voltage. Here we speak of the phase of the displacement x referred to the driving force F. The two phases are not equivalent because dx/dt, and not x, is the analog of the current.

Equation (112) can only be satisfied if the coefficient of cos ωt is zero. This condition may be written as

$$\tan\varphi = \frac{\sin\varphi}{\cos\varphi} = -\frac{\omega/\tau}{\omega_0{}^2 - \omega^2}. \tag{113}$$

It is also necessary that the coefficient of sin ωt be zero:

$$x_0 = \frac{\alpha_0}{(\omega_0{}^2 - \omega^2)\cos\varphi - (\omega/\tau)\sin\varphi}. \tag{114}$$

From (113) it follows that

$$\cos\varphi = \frac{\omega_0{}^2 - \omega^2}{[(\omega_0{}^2 - \omega^2)^2 + (\omega/\tau)^2]^{1/2}};$$

$$\sin\varphi = \frac{-\omega/\tau}{[(\omega_0{}^2 - \omega^2)^2 + (\omega/\tau)^2]^{1/2}}, \tag{115}$$

whence (114) becomes

$$\boxed{x_0 = \frac{\alpha_0}{[(\omega_0{}^2 - \omega^2)^2 + (\omega/\tau)^2]^{1/2}}.} \tag{116}$$

This is the amplitude of the motion.

Equations (115) and (116) give us the desired solution. We now know the amplitude x_0 and phase φ of the response of the system under the driving force $F = M\alpha_0 \sin\omega t$:

$$x = \frac{\alpha_0}{[(\omega_0{}^2 - \omega^2)^2 + (\omega/\tau)^2]^{1/2}} \sin\left(\omega t + \tan^{-1}\frac{\omega/\tau}{\omega^2 - \omega_0{}^2}\right). \tag{117}$$

We can develop a feeling for this solution by examining limiting cases. In our discussion we assume always that the damping is light, so that $\omega_0\tau \gg 1$.

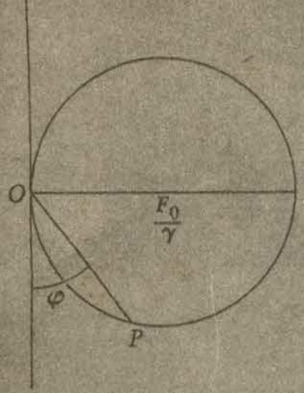
Low Driving Frequency, $\omega \ll \omega_0$. Here we see from (115) that

$$\cos\varphi \to 1; \qquad \sin\varphi \to -0, \tag{118}$$

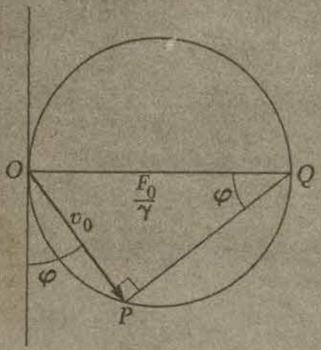
whence $\varphi \to 0$. This response at low frequency is said to be *in phase* with the driving force. From (116) we see that

$$x_0 \to \frac{\alpha_0}{\omega_0{}^2} = \frac{M\alpha_0}{C} = \frac{F_0}{C}. \tag{119}$$

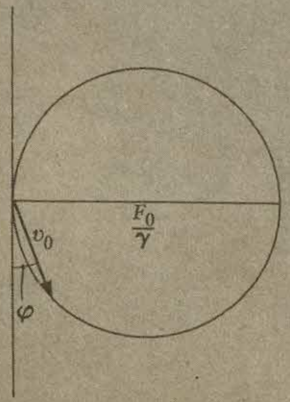
The *spring* (and not the mass or the friction) controls the response in this limit.



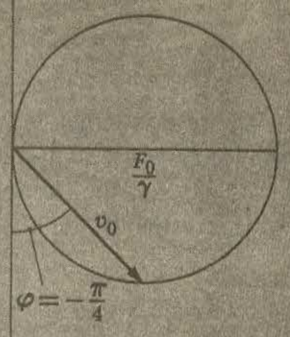
The "polar plot" gives a simple graphical representation of the driven harmonic oscillator. A circle with diameter F_0/γ is constructed, and a line segment OP making angle φ with the ordinate is drawn.



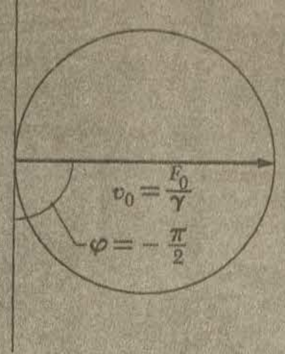
For any φ the triangle OPQ is a right triangle. Thus line segment $OP = -\frac{F_0}{\gamma}\sin\varphi$. From equations (115), (116), and (117) we see that line segment $OP = \omega x_0 = v_0$, the amplitude of the velocity.



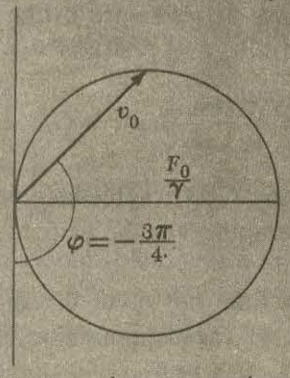
For $\omega \ll \omega_0$, $\varphi \simeq 0$ and $v_0 \ll F_0/\gamma$. The response is very small in this limit.



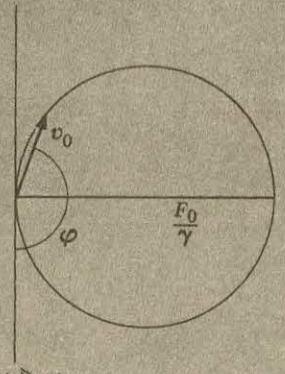
As ω grows $|\varphi|$ increases and so does v_0. At $\varphi = -\pi/4$, $v_0 = \dfrac{F_0}{\sqrt{2}\gamma}$.



At $\varphi = -\pi/2$, $\omega = \omega_0$ and $v_0 = \dfrac{F_0}{\gamma}$. The velocity amplitude is a *maximum at resonance.*

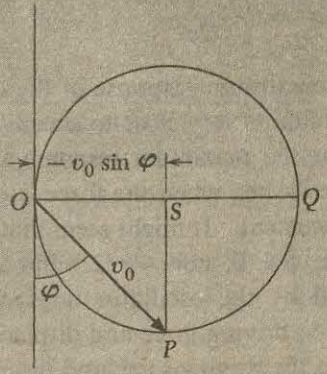


For $\omega > \omega_0$, v_0 decreases again. At $\varphi = -3\pi/4$, $v_0 = \dfrac{F_0}{\sqrt{2}\gamma}$ again.

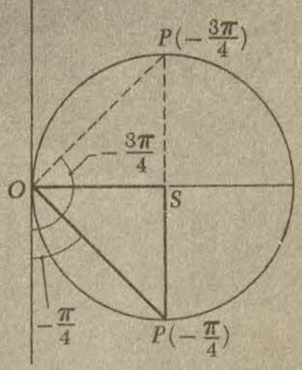


For $\omega \gg \omega_0$,

$$v_0 \ll \frac{F_0}{\gamma} \text{ again, } \quad \text{and} \quad \varphi \approx -\pi.$$



The line segment $OS = -v_0 \sin\varphi$ is shown. From equations (127), (128), and (129) we see that the power absorbed is proportional to $-v_0 \sin\varphi$, or to line segment OS.



At phase angles $\varphi = -\pi/4$ and $\varphi = -3\pi/4$, line segment $OS = \frac{1}{2}OS_{\max}$. Thus *half* the maximum power absorption occurs at these points. Of course *maximum* power absorption occurs at $\varphi = -\pi/2$ (resonance.)

Resonance Response, $\omega = \omega_0$. The response may be very large at resonance. We often use resonance response in applications, and we need to treat it carefully. At $\omega = \omega_0$ the driving frequency equals the natural frequency of the system in the absence of friction. We have

$$\cos\varphi \rightarrow \pm 0; \qquad \sin\varphi \rightarrow -1; \qquad \varphi \rightarrow -\frac{\pi}{2}. \tag{120}$$

The amplitude at $\omega = \omega_0$ is given by

$$x_0 = \frac{\alpha_0 \tau}{\omega_0}. \tag{121}$$

The lower the damping, the higher are τ and x_0. Keeping F_0 constant, the ratio of the response at resonance to the response at zero frequency is given from (119) and (120) as

$$\frac{x_0(\omega = \omega_0)}{x_0(\omega = 0)} = \frac{\alpha_0\tau/\omega_0}{\alpha_0/\omega_0{}^2} = \omega_0\tau = Q, \tag{122}$$

with the Q factor as defined by (103). This may be very large, often 10^4 or more! The damping can be said to control the response at resonance.

The maximum response x_0 does not occur at exactly $\omega = \omega_0$. We note that the derivative of the denominator of (116) has a zero at

$$\frac{d}{d\omega}\left[(\omega_0{}^2 - \omega^2)^2 + \left(\frac{\omega}{\tau}\right)^2\right] = 2(\omega_0{}^2 - \omega^2)(-2\omega) + \frac{2\omega}{\tau^2} = 0, \tag{123}$$

or

$$\omega^2 = \omega_0{}^2 - \frac{1}{2\tau^2}. \tag{124}$$

This is the position of the maximum response of the curve of x_0 vs ω. If $\omega_0\tau \gg 1$, the maximum is very close to $\omega = \omega_0$.

It might appear odd that the maximum response is obtained with the phase angle in $-\pi/2$, i.e., when the force is exactly 90° out of phase with the displacement. It might seem that resonance would logically occur when $\varphi = 0$, not $-\pi/2$. But here is the catch: the power absorbed by the oscillator does not depend directly on the phase between driving force and displacement, but rather on the phase between the driving force and the *velocity*. It takes about a minute's reflection to see that we will obtain the largest deflections when the velocity is exactly in phase with the driving forces. In this manner the mass gets pushed at just the right times and places. When the displacement is zero the velocity is greatest. If at this point it is moving in the positive direction, we wish the force at this time to attain its greatest value, in order to obtain maximum motion. At the turning points where the velocity changes its direction we wish, if we want resonance, the force to change its direction in the same manner at that moment. Thus resonance is best seen in terms of the phase between velocity and driving force. We know that the velocity of an oscillator leads its displacement by exactly 90°. Thus for resonance, with force and velocity in phase, we must have the force 90° ahead of the displacement, so that $\varphi = -\pi/2$.

High Driving Frequency, $\omega \gg \omega_0$. Here

$$\cos\varphi \to -1; \qquad \sin\varphi \to 0; \qquad \varphi \to -\pi, \tag{125}$$

and

$$x_0 \to \frac{\alpha_0}{\omega^2} = \frac{M\alpha_0}{M\omega^2} = \frac{F_0}{M\omega^2}. \tag{126}$$

In this limit the response decreases as $1/\omega^2$. The inertia of the mass controls the response in the high-frequency limit.

Notice that the phase φ of the displacement x with respect to the driving force F starts from zero at low frequencies, passes through $-\frac{1}{2}\pi$ at resonance, and attains $-\pi$ at high frequencies. *The displacement always lags the driving force.*

Power Absorption. The time average of the work done per unit time on the oscillating system by the driving force is given, using (106) and (117), by

$$P = \langle F\dot{x}\rangle = \frac{M\alpha_0{}^2\omega}{[(\omega_0{}^2 - \omega^2)^2 + (\omega/\tau)^2]^{1/2}}\langle \sin\omega t\cos(\omega t + \varphi)\rangle. \tag{127}$$

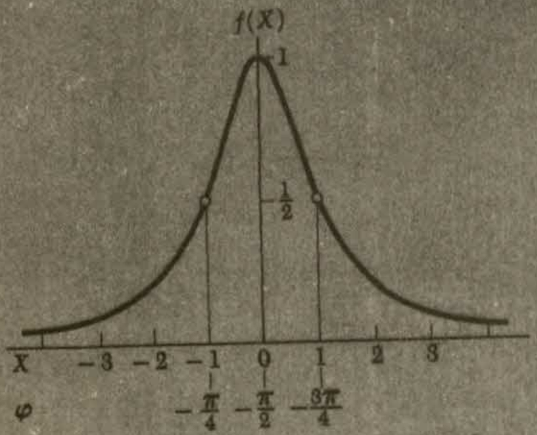
Using the identity

$$\cos(\omega t + \varphi) = \cos\omega t\cos\varphi - \sin\omega t\sin\varphi,$$

we have

$$\begin{aligned}\langle \sin\omega t\,[\cos\omega t\cos\varphi - \sin\omega t\sin\varphi]\rangle &= -\sin\varphi\,\langle \sin^2\omega t\rangle \\ &= -\tfrac{1}{2}\sin\varphi,\end{aligned} \tag{128}$$

where we have used the fact that $\langle \sin\omega t\cos\omega t\rangle = 0$. We see that



The power absorption is proportional to $f(X) = 1/(1 + X^2)$, where

$$X = \frac{-(\omega_0^2 - \omega^2)}{\omega/\tau} = \cot\varphi,$$

according to Eq. (139). For $\varphi \approx 0$, $X = \cot\varphi$ is large and negative. For $\varphi = -\pi/2$, $X = \cot\varphi = 0$ and the power absorption is a maximum. The half-power points $X = \pm 1$ are also seen. The function $f(X) = 1/(1 + X^2)$ is known as a Lorentz function.

the phase is important here. With (115) for sin φ we may write (127) as

$$\boxed{P = \frac{1}{2}M\alpha_0^2\frac{\omega^2/\tau}{(\omega_0^2 - \omega^2)^2 + (\omega/\tau)^2}.} \tag{129}$$

This is a very important result.

The resonance power absorption at $\omega = \omega_0$ is

$$P_{\text{res}} = \tfrac{1}{2}M\alpha_0^2\tau. \tag{130}$$

The power absorption (129) is reduced to one-half of the value at resonance when ω is changed by $\pm(\Delta\omega)_{1/2}$ so that

$$\frac{\omega}{\tau} = \omega_0^2 - \omega^2 \equiv (\omega_0 + \omega)(\omega_0 - \omega) \cong 2\omega_0(\Delta\omega)_{1/2}. \tag{131}$$

Thus the full width $2(\Delta\omega)_{1/2}$ of the resonance at half-maximum power is equal to $1/\tau$. We see, using the expression for Q found in (103), that

$$Q = \omega_0\tau = \frac{\omega_0}{2(\Delta\omega)_{1/2}} = \frac{\text{frequency at resonance}}{\text{full width at half-maximum power}}. \tag{132}$$

Thus Q measures the sharpness of tuning.

Example. Numerical example of a harmonic oscillator problem. Let the mass $M = 1$ gm; the force constant $C = 10^4$ dynes/cm; and the relaxation time $\tau = \frac{1}{2}$ sec. Then, from (82),

$$\omega_0 = \left(\frac{C}{M}\right)^{1/2} = 10^2\ \text{sec}^{-1},$$

while from (100) the free oscillation frequency is

$$\left[\omega_0^2 - \left(\frac{1}{2\tau}\right)^2\right]^{1/2} = [10^4 - 1]^{1/2} \cong 10^2\ \text{sec}^{-1}.$$

The Q of the system is given by (103) as

$$Q \cong \omega_0\tau = (10^2)(\tfrac{1}{2}) = 50.$$

The time for the amplitude to damp to e^{-1} of its initial value (for the free system) is

$$2\tau = 1\ \text{sec},$$

using (91). The damping constant $\gamma = M/\tau = 2$ gm/sec.

Now let the system be driven by the force

$$F = M\alpha_0 \sin \omega t = 10 \sin 90t \text{ dynes}.$$

We see that $\alpha_0 = F_0/M = 10$ dynes/gm, and the driving frequency is $\omega = 90 \text{ sec}^{-1}$. The amplitude is given by (116):

$$x_0 \cong \frac{10}{[4 \times 10^6 + 4 \times 10^4]^{1/2}} \cong 5 \times 10^{-3} \text{ cm},$$

and the phase is given by (113):

$$\tan \varphi \cong -\frac{180}{1.9 \times 10^3} \cong -0.1,$$

or $\varphi \cong -0.1 \text{ rad} \cong -6^\circ$. Thus in every cycle the maximum of the displacement occurs at 0.1 rad/90 rad per sec $\approx 10^{-3}$ sec after the maximum of the force.

We may compare the amplitude above with that in the limit $\omega \to 0$ and with that at resonance. From (119) we have $x_0(\omega = 0) = \alpha_0/\omega_0{}^2 = 10/10^4 = 10^{-3}$ cm. At resonance we have from (122) that

$$x_0(\omega = \omega_0) = Qx_0(\omega = 0) = (50)(10^{-3}) = 5 \times 10^{-2} \text{ cm}.$$

The full width of the resonance curve between half-power points is

$$2(\Delta\omega)_{1/2} = \frac{\omega_0}{Q} = \frac{100}{50} = 2 \text{ sec}^{-1}.$$

Note that in this example we have everywhere used "frequency" to mean "angular frequency." To get ordinary frequencies in oscillations per second or cycles per second we must divide by 2π.

Superposition Principle. An important property of the harmonic oscillator is that the solutions are additive: If $x_1(t)$ is the motion under the driving force $F_1(t)$, and $x_2(t)$ is the motion under the driving force $F_2(t)$, then $x_1(t) + x_2(t)$ is the motion under the combined force $F_1(t) + F_2(t)$. That is, if we know the motion x_1 under F_1 alone and the motion x_2 under F_2 alone, then we obtain the motion under the combined forces by adding together x_1 and

x_2. This property follows directly from the equation of motion:

$$\begin{aligned}\left(\frac{d^2}{dt^2}+\frac{1}{\tau}\frac{d}{dt}+\omega_0^2\right)&(x_1+x_2)\\ &=\left(\frac{d^2}{dt^2}+\frac{1}{\tau}\frac{d}{dt}+\omega_0^2\right)x_1+\left(\frac{d^2}{dt^2}+\frac{1}{\tau}\frac{d}{dt}+\omega_0^2\right)x_2\\ &=F_1+F_2. \end{aligned} \tag{133}$$

The validity of the superposition principle for the solutions of the harmonic oscillator equation of motion is a consequence of the linearity of the equation; only the first power of x enters. The picture is quite different when anharmonic terms are included. A term in x^2 in the equation of motion can be shown to mix and to multiply two simultaneous driving frequencies ω_1 and ω_2, thereby producing a full range of harmonic frequencies ($2\omega_1, 3\omega_1, \ldots, 2\omega_2, 3\omega_2, \ldots$) and of combination or "sideband" frequencies ($\omega_1+\omega_2$; $\omega_1-\omega_2$; $\omega_1-2\omega_2$; etc.).

Further reading

PSSC, chaps. 16 (paragraph 3), 21 (paragraph 8), 24 (paragraph 10), 25 (paragraph 1), 26 (paragraphs 6 and 7), and 31 (paragraph 9).

Y Rocard, *General dynamics of vibrations* (Ungar, New York, 1960). A simple and lucid book written with a broad range of applications in mind.

B. L. Walsh, "Parametric amplification," *International Science and Technology*, no. 17, p. 75 (May 1963). An elementary discussion of parametric amplifiers and their property of low noise.

Film list

"Periodic Motion" (33 min) Hume and Ivey (PSSC-MLA 0306). An excellent film on simple harmonic motion.

Problems

1. *Simple pendulum.* A pendulum is constructed from a light string of length $L = 100$ cm and a heavy mass $M = 1 \times 10^3$ gm.

(*a*) What is the period of the pendulum for small displacements?

Ans. 2.0 sec.

(*b*) What would be the approximate change in this value if the initial displacement of the mass was 60°?

Ans. New period is 2.1 sec.

2. *Mass on a spring.* A mass of 1×10^3 gm is suspended from a linear spring with a force constant $C = 1 \times 10^6$ dynes/cm and a damping coefficient γ of 50 dyne-sec/cm. The spring is driven by an external force

$F = F_0 \sin \omega t$, where $F_0 = 2.5 \times 10^5$ dynes and ω is twice the natural frequency of the system. What is the amplitude of the resulting motion? How much (in radians) is the displacement shifted in phase from the driving force? Remember that α_0 in the text was defined as F_0/M.

Ans. 8.3×10^{-2} cm; $-179.9°$.

3. *Mass on a spring—data.* The data below were obtained by observing the motion of a mass attached to the end of a spring:

Table 1. Period as a function of mass

Mass, gm	*Observed periods, sec*
50	0.72
100	0.85
150	0.96
200	1.06
250	1.16
300	1.23

Table 2. Amplitude of vibration as a function of time for mass 150 gm

Time, sec	*Amplitude, cm*
0	4.5
30	4.0
80	3.5
125	3.0
180	2.5
235	2.0
340	1.5
455	1.0

(*a*) Plot the square of the period of oscillation as a function of mass. The tabulated values of mass exclude the mass of the spring. Determine the effective mass of the spring by proper extrapolation of the graph.

(*b*) Determine the spring constant C.

(*c*) Plot the natural logarithm of the amplitude as a function of time and determine the relaxation time.

(*d*) Determine the damping factor γ.

4. *Mass on a spring—energy.* Consider the data in Table 2 in Prob. 3.

(*a*) What is the average total energy at $t = 100$ sec?

(*b*) What is the average rate of energy loss at $t = 100$ sec?

(*c*) What is the energy loss in one period at $t = 100$ sec? in 1 rad of the motion?

(*d*) What is the Q of the system?

5. *Mass on a spring—time intervals.* Consider the motion of a mass $M = 5$ gm on the end of a vertical spring of force constant $C = 20$ dynes/cm. The initial position and velocity are 5 cm above the equilibrium point and 2 cm/sec downward.

(*a*) Show that the change in velocity of the particle, for the first interval following release, is

$$\Delta v_1 = -\frac{C}{M} x_0 \, \Delta t.$$

(*b*) Show that the corresponding change in position is

$$\Delta x_1 = v_0 \, \Delta t.$$

(*c*) Choose successive time intervals of 0.2 sec and repeat the process indicated in parts (*a*) and (*b*) to obtain position and velocity as a function of time for 4 sec. It will be convenient to display the results in a table:

t, sec	x, cm	Δv, cm/sec	v, cm/sec	Δx, cm
0	5.00	−4.00	−2.00	−0.40
0.2	4.60	−3.68	−6.00	−1.20
0.4	3.40	—	—	—
—	—	—	—	—
—	—	—	—	—

When you are through, make an accurate graph of $x(t)$.

(*d*) The solution may be written

$$x(t) = \frac{v_0}{\omega} \sin \omega t + x_0 \cos \omega t.$$

For the particular conditions of this problem the solution is

$$x(t) = \sin 2t + 5 \cos 2t.$$

Plot this result on the same graph as the numerical calculation of part (*c*).

6. *Differential equation for simple pendulum.* Set up the differential equation for the problem of the simple pendulum by using $\mathbf{F} = M\mathbf{a}$ directly. Work with the component of the gravitational force perpendicular to the pendulum rod, when the rod is at an angle θ.

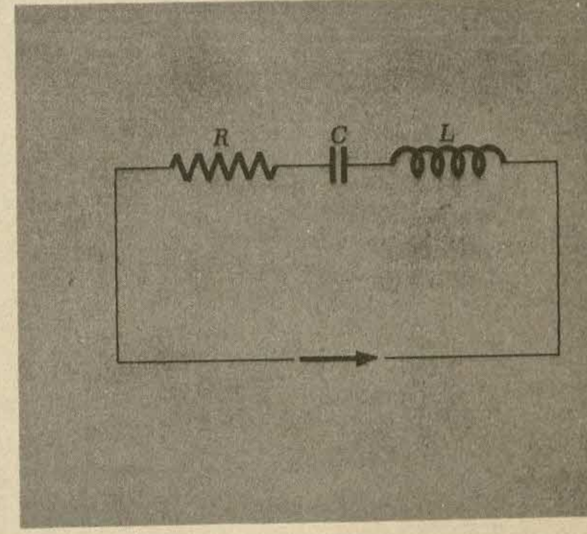
7. *Velocity of object traveling in viscous medium.* An object falling through a viscous medium experiences a retarding force $-\gamma v$. For example, in the Millikan oil-drop experiment a charged oil drop of mass M and charge q falls under the combined influences of an electric field E and of gravity g. The particle quickly reaches a terminal velocity v_{ter}. Set up and solve the equation of motion to obtain the velocity of the particle as a function of time. [*Hint:* Assume a solution of the form $v = A + Be^{-\alpha t}$ and determine α, A, B from the equation and the values of v at $t = 0$ and at $t = \infty$.] By considering the limit $t \rightarrow \infty$, show that the terminal velocity is

$$v_{\text{ter}} = \left(\frac{q}{M}\tau\right)E + g\tau,$$

where $\tau \equiv M/\gamma \equiv$ relaxation time. Observations of the terminal velocity as a function of the electric field provide a good means of determining the relaxation time τ and, hence, the damping factor γ. In a typical experiment two parallel plates separated by 0.7 cm are maintained at a potential difference of 840 volts. [The electric field in electrostatic units is (840/0.7)/300 statvolts/cm.] If the terminal velocity of a singly charged particle of mass 2×10^{-12} gm is 0.01 cm/sec, calculate the relaxation time τ. Here take the force due to the electric field to be in the same direction as the force due to gravity. *Ans.* $\tau = 5 \times 10^{-6}$ sec.

8. *Simple series LCR circuit.* A simple series circuit containing a resistor R, a capacitor C, and an inductor L obeys the same differential equation as the freely oscillating damped spring. The equation for the current I is

$$L\frac{d^2I}{dt^2} + R\frac{dI}{dt} + \frac{1}{C}I = 0.$$



By analogy with the spring, find expressions in terms of L, R, C for the relaxation time τ and for the natural frequency ω_0.

9. *Driven LCR circuit.* Suppose that the circuit of Prob. 8 is driven by an alternating voltage source $V(t) = V_0 \sin \omega t$. The differential equation becomes

$$L\frac{d^2I}{dt^2} + R\frac{dI}{dt} + \frac{1}{C}I = \omega V_0 \cos \omega t.$$

Look for a solution of the form $I = I_0 \cos (\omega t + \varphi)$ and solve for I_0 and φ, in terms of X_L, X_C, and R, where $X_L = \omega L$ and $X_C = 1/\omega C$.

10. *Damped driven harmonic oscillator.* Discuss the limiting cases below of a damped driven harmonic oscillator. In each instance give in words a *physical explanation* of what you expect the behavior of the system to be and compare with the appropriate limit of the solution (117) and (129) of the complete problem.

(*a*) $M \rightarrow 0$, other quantities F_0, ω, τ, C finite.

(*b*) $C \rightarrow 0$, other quantities finite.

(*c*) $\tau \to \infty$, other quantities finite.

(*d*) $M \to 0$; $C \to 0$, other quantities finite.

(*e*) $M \to 0$; $\tau \to \infty$, other quantities finite.

11. *Measurement of earth's gravitational field.* The strength of the earth's gravitational field can be measured by determining the period of a precision pendulum. This apparatus may also be used to determine the acceleration of an object in the vertical plane. For example, at a point for which $g = 980\ \text{cm/sec}^2$, the length of the pendulum is adjusted so that the period is 1 sec. When the period of the pendulum is measured on an elevator undergoing uniform acceleration, it is found to be 1.025 sec.

(*a*) In the limit of an acceleration a, small compared to gravity, show that the period measured on the elevator is given by

$$T \cong T_0\left(1 - \frac{1}{2}\frac{a}{g}\right),$$

where an upward acceleration is positive. Here T_0 is the period of the unaccelerated pendulum.

(*b*) What is the acceleration of the elevator?

12. *Torsion suspension.* The torsion suspension has been used in many instruments. Coulomb and Cavendish used it in their famous experiments. It is still used in several types of electrometers and magnetometers, and in the Wood-Anderson torsion seismometer. The torque N of the suspending fiber is proportional to the angle of twist φ, so that $N = -K\varphi$. The fiber carries an object whose moment of inertia is I, where I relates the angular momentum $\mathbf{J}$ to the angular velocity ω by $\mathbf{J} = I\omega$, as defined in Chap. 8.

(*a*) Show that the equation of motion for small φ can be written in the same form as for a mass on a spring.

(*b*) Solve the equation of motion for the frequency of torsional oscillation.

13. *Dimensionless form for oscillator response.* The amplitude of oscillation and the phase shift depend upon the frequency of the driving force. These expressions may easily be written in a dimensionless form, applicable to *all* driven harmonic oscillators. Show that

$$\tan\varphi = \frac{\theta/\theta_0}{\theta^2 - 1} \qquad \text{and} \qquad \frac{\omega_0{}^2 x_0}{\alpha_0} = \frac{1}{\left[(1-\theta^2)^2 + \left(\dfrac{\theta}{\theta_0}\right)^2\right]^{1/2}},$$

where $\theta = \omega/\omega_0$ and $\theta_0 = \omega_0\tau$.

14. *Quality factor Q.* By definition the quality factor Q for a driven harmonic oscillator is

$$Q = \frac{2\pi\langle\text{energy stored}\rangle}{\langle\text{energy lost per cycle}\rangle} = \frac{\langle E\rangle\omega}{P}.$$

(*a*) Show for the driven oscillator that the average total energy is

$$\langle E\rangle = \langle K + U\rangle = \langle K\rangle + \langle U\rangle = \tfrac{1}{4}M\omega^2 x_0^2 + \tfrac{1}{4}M\omega_0^2 x_0^2.$$

(*b*) Use these expressions to show

$$Q = \frac{1}{2}\left[1 + \left(\frac{\omega_0}{\omega}\right)^2\right](\omega\tau).$$

15. *Mass on a spring—gravity.* Consider a mass M suspended from a vertical spring of force constant C. Show that the force of gravity acting on M does not change the period of oscillation but only the center point of the oscillation.

16. *Complex variables.* Consider the solution of (82) using the notation of complex variables. Look for a solution of the form $x = x_0 e^{i\omega t}$, where ω may be complex. Show that

$$\omega = \left[\omega_0^2 - \left(\frac{1}{2\tau}\right)^2\right]^{1/2} + \frac{i}{2\tau}.$$

Note that if $\omega = \omega_R + i\omega_I$, then

$$e^{i\omega t} = e^{-\omega_I t} e^{i\omega_R t},$$

so that the solution is a damped oscillation.

17. *Viscosity.* According to Stokes' law (which is valid at sufficiently low velocities) the damping coefficient of a sphere of radius R in a fluid of viscosity η is given by $\gamma = 6\pi R\eta$.

(*a*) Consult a reference book to give and explain the definition of *viscosity*.

(*b*) What are the dimensions of the viscosity?

(*c*) What is the name of the CGS unit of viscosity?

(*d*) What is the value of the viscosity of water at 30°C?

(*e*) Evaluate γ for a sphere of radius 10 cm in a medium of viscosity 2 centipoise. *Ans.* $\simeq 4$ gm/sec.

Advanced Topic 1. Exact solution, problem of the simple pendulum

The energy of the simple pendulum is

$$E = \frac{1}{2}ML^2\dot{\theta}^2 + MgL(1 - \cos\theta), \tag{134}$$

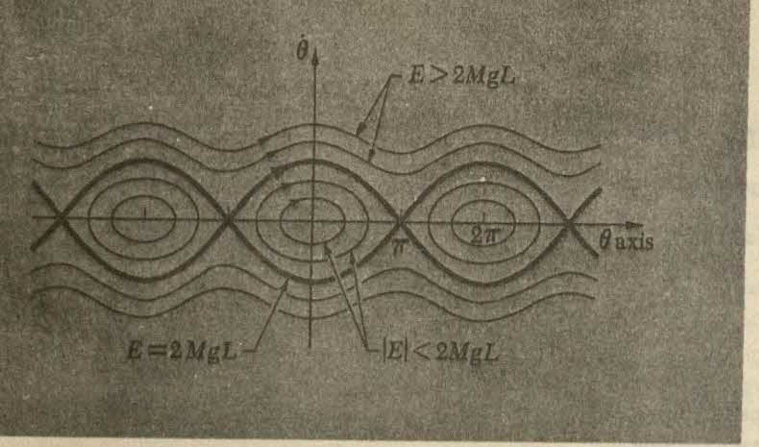
or

$$\dot{\theta}^2 = 2\omega_0^2\cos\theta + C, \tag{135}$$

where

$$\omega_0^2 = \frac{g}{L}; \qquad C = \left(\frac{2E}{ML^2} - \frac{2g}{L}\right). \tag{136}$$

Suppose the pendulum is let fall from rest at an initial displacement $-\pi < \theta_0 < \pi$. Disregarding friction, we expect the motion to be periodic



but not simple harmonic, with $\dot\theta = 0$ at $\theta = \pm\theta_0$. If the pendulum is set in motion by a sufficiently large push, however, it will continue to rotate in one direction. The motion will repeat itself periodically, but $\dot\theta$ will never be 0, and $\theta(t)$ will continue to increase.

These situations are clearly distinguished when we examine the *phase plane* of the pendulum's motion, in which the motion is described by plotting θ against $\dot\theta$. The phase plane plays an important role in the solution of nonlinear differential equations. Clearly, for fixed values of ω_0^2 and C, Eq. (135) describes a curve in the phase plane. For different values of these constants we obtain the family of curves sketched in the figure. The *closed* curves about the point $\theta = 0$, $\dot\theta = 0$ correspond to values of C and ω_0^2 permitting periodic motion of the pendulum back-and-forth through the equilibrium point; the open curves correspond to motions for which the initial values of these constants are such that the pendulum rotates in one direction. The duplicate families of closed curves about the points $(2n\pi,0)$ have no additional physical significance. From physical considerations, $C \geqq -2\omega_0^2$, because the total energy cannot be negative. If $E = 0$, then $C = -2\omega_0^2$; in this case the pendulum remains at rest, and its path in the phase plane is the point (0,0). The closed curves correspond to values of C for which $-2\omega_0^2 < C \leqq 2\omega_0^2$, as may be seen from the physical meaning of this requirement: namely, that $2MgL \geqq E$.

We may find the period and frequency of a motion described by a closed curve in the phase plane. The maximum amplitude θ_0 will be attained when $\dot\theta = 0$; thus

$$\cos\theta_0 = -\frac{C}{2\omega_0^2}. \tag{137}$$

The period T is four times the time required to swing from $\theta = 0$ to $\theta = \theta_0$. From (135) we have

$$\frac{d\theta}{dt} = \dot\theta = (2\omega_0^2\cos\theta + C)^{1/2}; \tag{138}$$

the period is

$$\begin{aligned} T &= 4\int_0^{\theta_0} \frac{d\theta}{(2\omega_0^2\cos\theta + C)^{1/2}} \\ &= \frac{4}{(2\omega_0^2)^{1/2}} \int_0^{\theta_0} \frac{d\theta}{(\cos\theta - \cos\theta_0)^{1/2}}. \end{aligned} \tag{139}$$

We transform this integral into a standard form by introducing a new variable ψ (Greek psi), defined by:

$$\sin\psi = \frac{\sin(\theta/2)}{\sin(\theta_0/2)}. \tag{140}$$

(*a*) Show that

$$T = \frac{4}{\omega_0}\int_0^{\pi/2} \frac{d\psi}{[1 - \sin^2(\theta_0/2)\sin^2\psi]^{1/2}} = \frac{2\pi}{\omega}. \tag{141}$$

The integral

$$\int_0^{\pi/2} \frac{d\psi}{(1 - K^2\sin^2\psi)^{1/2}}$$

is called a *complete elliptic integral of the first kind;* extensive tables exist for its values (corresponding to different values of K^2); see, for example, Dwight 773.1 and 1040. Here K is just a constant.

The solution of the elliptic integral in series form is:

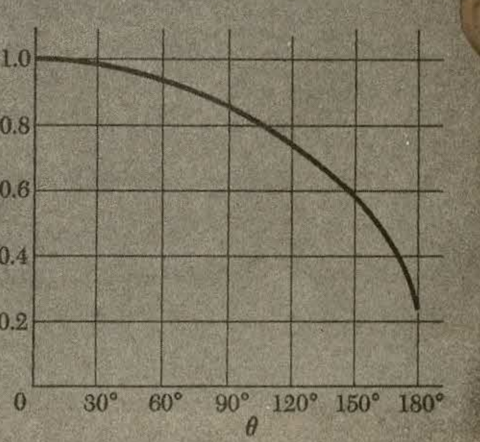
$$\int_0^{\pi/2} \frac{d\psi}{(1 - K^2\sin^2\psi)^{1/2}} = \frac{\pi}{2}\left\{1 + \sum_{n=1}^{\infty}\left[\frac{1\cdot 3\cdot 5\cdot \cdots \cdot (2n-1)}{2\cdot 4\cdot 6\cdot \cdots \cdot (2n)}\right]^2 K^{2n}\right\}. \tag{142}$$

If we write out the first few terms of (142),

$$\begin{aligned} \omega &= \omega_0\left[1 + \frac{1}{4}\sin^2\left(\frac{\theta_0}{2}\right) + \frac{9}{64}\sin^4\left(\frac{\theta_0}{2}\right) + \cdots\right]^{-1} \\ &\cong \omega_0\left(1 - \frac{1}{16}\theta_0^2 + \cdots\right), \end{aligned} \tag{143}$$

in agreement with (38).

(*b*) Show that $\omega = 0.847\ \omega_0$ at $\theta_0 = \pi/2$.



Advanced Topic 2. Anharmonic oscillator

There is no physical reason why a real spring should not contain a term in x^2 (or $x^3 \ldots$) in the force law, which means x^3 (or $x^4 \ldots$) in the potential energy. The potential energy function of a real spring need not be exactly symmetrical about the equilibrium position. If the potential energy of the spring is

$$V(x) = \tfrac{1}{2}Cx^2 - \tfrac{1}{3}sCx^3, \tag{144}$$

where s may be called the anharmonicity constant, then the force due to the spring is†

$$F_x = -\frac{dU}{dx} = -Cx + sCx^2. \tag{145}$$

---

† The force here is zero both for $x = 0$ and for $x = 1/s$. We assume that the amplitude of the motion is small in comparison with $1/s$, so that the particle stays near the minimum at $x = 0$ of the potential energy (144).

The minus sign in (144) was introduced to correspond to a common physical situation in which the force gets weaker for large positive x. The equation of motion is nonlinear:

$$M\ddot{x} + Cx - sCx^2 = 0; \qquad \ddot{x} + \omega_0^2 x - s\omega_0^2 x^2 = 0. \tag{146}$$

The x^2 term makes the equation nonlinear.

Now we look for a solution of the form

$$x = A(\cos\omega t + q\cos 2\omega t) + x_1, \tag{147}$$

where q and x_1 are constants to be determined. The selection of $\cos\omega t$ is a matter of convenience, for we could have used any linear combination of $\cos\omega t$ and $\sin\omega t$. But having selected $\cos\omega t$, the mathematics forces us to use $\cos 2\omega t$ also. (This is found by trial and error.)

We form

$$\ddot{x} = -\omega^2 A(\cos\omega t + 4q\cos 2\omega t); \tag{148}$$

$$x^2 = A^2(\cos^2\omega t + \cdots) = \tfrac{1}{2}A^2 + \tfrac{1}{2}A^2\cos 2\omega t + \cdots, \tag{149}$$

using the identity $\cos^2\omega t = \frac{1}{2}(1 + \cos 2\omega t)$. In x^2 we neglect terms in q and in x_1, because x^2 enters the equation of motion multiplied by s. We assume $sA \ll 1$. The differential equation (146) becomes

$$-\omega^2 A(\cos\omega t + 4q\cos 2\omega t) + \omega_0^2 A(\cos\omega t + q\cos 2\omega t) + \omega_0^2 x_1 - \tfrac{1}{2}s\omega_0^2 A^2 - \tfrac{1}{2}s\omega_0^2 A^2\cos 2\omega t + \cdots = 0. \tag{150}$$

The condition that the coefficient of $\cos\omega t$ in (150) should vanish is

$$-\omega^2 + \omega_0^2 = 0, \tag{151}$$

so in this approximation there is no frequency shift. The condition that the coefficient of $\cos 2\omega t$ should vanish is

$$-3q - \tfrac{1}{2}sA = 0; \qquad q = -\tfrac{1}{6}sA. \tag{152}$$

The displacement x_1 is given from (150) by

$$x_1 - \tfrac{1}{2}sA^2 = 0; \qquad x_1 = \tfrac{1}{2}sA^2. \tag{153}$$

The time-average position is, from (147),

$$\langle x\rangle = A(\langle\cos\omega t\rangle + q\langle\cos 2\omega t\rangle) + x_1, \tag{154}$$

but $\langle\cos\omega t\rangle = 0$ and $\langle\cos 2\omega t\rangle = 0$, so that

$$\langle x\rangle = x_1 = \tfrac{1}{2}sA^2. \tag{155}$$

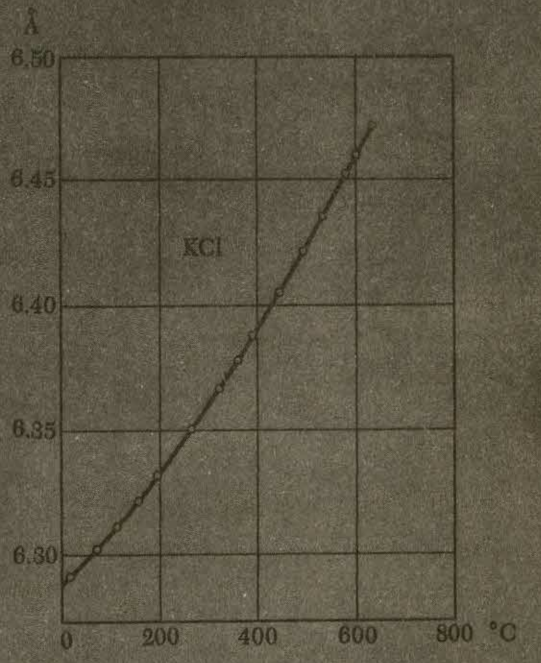
In the electrical analog of this problem, the application of an alternating potential produces a dc rectified voltage component in the response of the nonlinear circuit. What types of nonlinear electrical circuit elements can you imagine?

From (145) we see that the restoring force is stronger for x negative than for x positive. It is not surprising therefore to find a shift such as

(155) in the average position of the oscillating particle toward positive x as this side is softer. The shift (155) is proportional to the anharmonicity constant s and to the square of the oscillation amplitude. We know from our earlier results that the energy of a harmonic oscillator is proportional to A^2. It is a result of statistical physics (Vol. V) that the mean energy of a classical harmonic oscillator in thermal equilibrium is equal† to kT, where k is the Boltzmann constant and T is the absolute temperature. If this is true, we will have approximately that

$$\langle x \rangle \propto A^2 \propto \text{temperature}. \tag{156}$$

In this result we have the essential explanation of the thermal expansion of solids.‡ Experimental results for a potassium chloride crystal are given in the figure.



Length of unit cube of a crystal of potassium chloride (KCl) as a function of temperature in deg C. *(After R. E. Glover)*

Advanced Topic 3. Modulation of the oscillator parameters (parametric amplification)

We consider now a remarkable and important phenomenon, *parametric amplification*, or the generation of subharmonics by the modulation of one of the physical parameters of an oscillator. For example, in an RLC circuit the physical parameters of the oscillator are R, L, and C. We can modulate (change) the value of any of these quantities. Suppose in an undriven circuit we vary or modulate a capacitance by varying the separation of the plates.§ It is of particular interest to perform the modulation harmonically at the frequency $2\omega_0$, where ω_0 is the frequency of the original oscillator. We suppose that the damping is small.

We write the equation of the oscillator as

$$\ddot{x} + \frac{1}{\tau}\dot{x} + \omega_0^2[1 - \epsilon \sin 2\omega_0 t]x = 0, \tag{157}$$

where $\epsilon \ll 1$. The term in ϵ represents the modulation at frequency $2\omega_0$ of the restoring force of the oscillator. Let us look for solutions at the frequency ω_0, which is the resonant frequency of the circuit and is equal to one-half the modulation frequency:

---

† The result that the thermal energy is proportional to temperature is valid for almost all solids at room temperature and above, but fails at low temperatures because of quantum effects discussed in Vol. IV.

‡ The form of the potential energy between two atoms usually is strongly repulsive for x negative, so that s is positive and $\langle x \rangle$ is positive, corresponding to expansion as the solid is heated. The few instances known of thermal contraction in solids are chiefly due to effects of magnetic order of electron spins. In low-expansivity alloys such as Invar the effects of lattice expansion and magnetic contraction are made to balance each other over a useful range of temperature.

§ This is a legitimate but crude method; in microwave receivers the capacitance across a semiconductor diode would be varied by changing the voltage applied across the diode. The diode modulator, which is called a varactor, is widely used. Modulation of the capacitance is analogous to the modulation of the stiffness of a spring. In principle the stiffness could be modulated by alternately heating and cooling the spring, provided the value of C depended on the temperature.

$$x = e^{\beta t} \sin \omega_0 t. \tag{158}$$

The exponential term $e^{\beta t}$ is included because for zero modulation the amplitude will decrease; and we anticipate that for a sufficiently strong modulation amplitude ϵ there may be exponentially increasing solutions. That is, the modulation may set the system into even greater oscillation. On substituting (158) in (157) we find for the separate terms:

$$\begin{aligned} \ddot{x} &= \omega_0{}^2 e^{\beta t} \sin \omega_0 t + 2\beta\omega_0 e^{\beta t} \cos \omega_0 t + \beta^2 e^{\beta t} \sin \omega_0 t; \\ \frac{1}{\tau}\dot{x} &= \frac{\beta}{\tau} e^{\beta t} \sin \omega_0 t + \frac{\omega_0}{\tau} e^{\beta t} \cos \omega_0 t; \\ \omega_0{}^2 x &= \omega_0{}^2 e^{\beta t} \sin \omega_0 t; \\ -\epsilon\omega_0{}^2 x \sin 2\omega_0 t &= -\frac{1}{2}\epsilon\omega_0{}^2 e^{\beta t}(\cos \omega_0 t - \cos 3\omega_0 t). \end{aligned} \tag{159}$$

In the last line we have used the identity

$$2(\sin 2\omega_0 t) \sin \omega_0 t = \cos \omega_0 t - \cos 3\omega_0 t. \tag{160}$$

Now in (159) we suppose that the term of $O(\beta^2)$ is small in comparison with terms of $O(\beta\omega_0)$ and $O(\omega_0{}^2)$, which implies that $\beta \ll \omega_0$. This is consistent with the ultimate solution. Then the coefficient of the terms in $\sin \omega_0 t$ vanishes. The condition that the terms in $\cos \omega_0 t$ vanish is that

$$2\beta\omega_0 + \frac{\omega_0}{\tau} - \frac{1}{2}\epsilon\omega_0{}^2 = 0, \tag{161}$$

or

$$\beta = \frac{1}{4}\epsilon\omega_0 - \frac{1}{2\tau}. \tag{162}$$

Thus β will be positive and the amplitude of x will increase exponentially with time if

$$\epsilon > \frac{2}{\omega_0\tau}. \tag{163}$$

This is the condition for an oscillation of the circuit at the subharmonic ω_0 of the modulation frequency $2\omega_0$. We have *not* shown that ω_0 will be the only frequency observed in the system.

Note that below Eq. (160) we inserted the requirement that $\beta \ll \omega_0$. From Eq. (162) this yields

$$\epsilon \ll 4 + \frac{2}{\omega_0\tau}. \tag{164}$$

It is clear that we can choose ϵ to satisfy (164) *and* (163) simultaneously if $2/\omega_0\tau \ll 4$; i.e., if $Q = \omega_0\tau \gg \frac{1}{2}$.

What about the term in cos $3\omega_0 t$? Within our present approximation we have no way of taking this term into account. We are not even sure from this treatment how important the term may be. It is possible to solve (157) exactly in terms of tabulated functions called Mathieu functions, and we find results consistent with our approximate treatment. (The theory of Mathieu functions is not entirely elementary.) A simpler way to verify that parametric amplification does indeed occur is to replace the sinusoidal modulation with a square-wave modulation. This is left to the student. First replace (157) with

$$\ddot{x} + \frac{1}{\tau}\dot{x} + \omega_0^2(1 \pm \epsilon)x = 0, \tag{165}$$

where $-\epsilon$ applies for a time interval 0 to $\pi/2\omega_0$ and $+\epsilon$ applies from $\pi/2\omega_0$ to π/ω_0, and so on. Within each time interval the equation of motion (165) represents a simple harmonic oscillator, and the equation may easily be solved exactly. The natural frequencies are of course different in the two intervals. You obtain the entire exact solution by arranging to satisfy appropriate boundary conditions at the terminal points of the intervals. The boundary conditions are that x and $\dot{x}$ should be continuous everywhere.

INSTITUTE OF EDUCATION

Mathematical Note. Complex numbers and the driven harmonic oscillator

We now give a very neat solution of the problem of the driven harmonic oscillator, using the complex number scheme developed at the end of Chap. 4. The equation of motion (105) is (with cos ωt written for convenience in place of sin ωt),

$$\ddot{x} + \frac{1}{\tau}\dot{x} + \omega_0^2 x = \alpha_0 \cos \omega t. \tag{166}$$

Let us replace the force term by

$$\alpha_0 e^{i\omega t} \equiv \alpha_0 (\cos \omega t + i \sin \omega t). \tag{167}$$

At the end of the calculation we may take the response as the real part of x if the driving force is $\alpha_0 \cos \omega t$ (with α_0 real).

We look for a solution to (166) of the form

$$x = X_0 e^{i\omega t}, \tag{168}$$

where X_0 may be a complex number. On substitution of (168) in (166) we have

$$\left(-\omega^2 - \frac{i\omega}{\tau} + \omega_0^2\right) X_0 e^{i\omega t} = \alpha_0 e^{i\omega t}, \tag{169}$$

whence

$$X_0 = \frac{\alpha_0}{\omega_0^2 - \omega^2 + i(\omega/\tau)}. \tag{170}$$

It is useful to consider separately the real and imaginary parts of X_0. We have

$$\begin{aligned} X_0 &= \frac{\alpha_0}{\omega_0^2 - \omega^2 + i(\omega/\tau)} \cdot \frac{\omega_0^2 - \omega^2 - i(\omega/\tau)}{\omega_0^2 - \omega^2 - i(\omega/\tau)} \\ &= \alpha_0 \frac{\omega_0^2 - \omega^2 - i(\omega/\tau)}{(\omega_0^2 - \omega^2)^2 + (\omega/\tau)^2}, \end{aligned} \tag{171}$$

whence

$$\begin{aligned} \operatorname{Re}(X_0) &= \frac{(\omega_0^2 - \omega^2)\alpha_0}{(\omega_0^2 - \omega^2)^2 + (\omega/\tau)^2}; \\ \operatorname{Im}(X_0) &= \frac{-(\omega/\tau)\alpha_0}{(\omega_0^2 - \omega^2)^2 + (\omega/\tau)^2}. \end{aligned} \tag{172}$$

In the limit $|\omega_0^2 - \omega^2| \gg \omega/\tau$ we have

$$\operatorname{Re}(X_0) \approx \frac{\alpha_0}{\omega_0^2 - \omega^2}; \qquad \operatorname{Im}(X_0) \approx 0. \tag{173}$$

This condition is called *off-resonance,* and here the real part of X_0 is much more important than the imaginary part.

In the limit $|\omega_0^2 - \omega^2| \ll \omega/\tau$ we say we are *near resonance,* and for $\omega_0 = \omega$ we are *on the resonance* or on the center of the resonance. For $\omega_0 = \omega$,

$$\begin{aligned} \operatorname{Re}(X_0) &= 0; \\ \operatorname{Im}(X_0) &= \alpha_0 \frac{\tau}{\omega}. \end{aligned} \tag{174}$$

The greater τ is, the weaker the damping and the greater the imaginary part of the response at resonance.

Let us write X_0 in the form $\rho e^{i\varphi}$, as at the end of Chap. 4. Then from (170) we have for the amplitude of the response

$$\rho = (X_0 X_0^*)^{1/2} = \frac{\alpha_0}{[(\omega_0^2 - \omega^2)^2 + (\omega/\tau)^2]^{1/2}}. \tag{175}$$

Here X_0^* is the complex conjugate to X_0, so that $X_0 X_0^*$ is real. From (171) we have for the phase angle of x relative to F,

$$\tan \varphi = -\frac{\omega/\tau}{\omega_0^2 - \omega^2}. \tag{176}$$

The average power absorption is given by

$$\begin{aligned}\langle P\rangle &= \langle F\dot{x}\rangle = \langle \mathrm{Re}(F)\mathrm{Re}(\dot{x})\rangle \\ &= \langle [M\alpha_0 \cos \omega t][-\rho\omega \sin (\omega t + \varphi)]\rangle. \qquad (177)\end{aligned}$$

We have taken the real part of x to correspond to physical reality if the real part of F is the physically real force. There are other valid forms for the time average—what turns out to be important is to take that part of x which is in phase with F. Using (172) and the relation $\rho \sin \varphi = \mathrm{Im}(X_0)$, we have from (177)

$$\begin{aligned}\langle P\rangle &= -M\alpha_0\rho\omega \langle \cos^2 \omega t\rangle \sin \varphi = -\frac{1}{2}M\alpha_0\omega\,\mathrm{Im}(X_0) \\ &= \frac{1}{2}M\alpha_0{}^2\frac{\omega^2/\tau}{(\omega_0{}^2 - \omega^2)^2 + (\omega/\tau)^2}. \qquad (178)\end{aligned}$$

The result (178) is identical with the earlier result (129).

Chapter 8

Equations of motion of rotating body 238
  Inertial coefficients 240
  Examples: Showing **J** *not parallel to ω* 242
    Moment of inertia of spherical shell 243
    Moment of inertia of sphere 243
    Parallel-axis theorem 244
  Summation convention 245
Kinetic energy of rotation 246
  Example: Rolling sphere 248
  Principal axes 249
Euler equations 250
  Examples: Precession of sphere, no torque acting 251
    Symmetrical top 251
Spin precession in a constant magnetic field 253
  Spin resonance 254
Elementary gyroscope 255

Film list 257

Problems 258

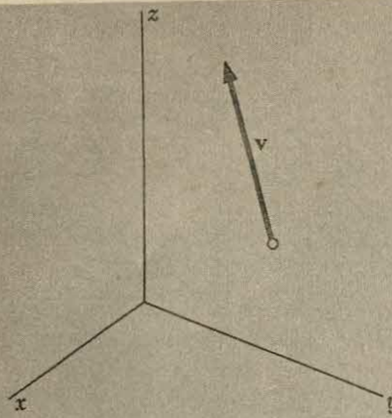
Elementary Dynamics of Rigid Bodies

The dynamics of rigid bodies is a fascinating and complicated subject, perhaps the highest point of classical mechanics and the most difficult. (*This chapter may be omitted on a first reading of this volume.*†) The problem is complicated because in general we need to find the solution of three simultaneous differential equations in the components of the angular velocity. We must consider the results both in a reference frame fixed in the rotating body and in an inertial reference frame in which the center of mass of the body is at rest. The prototype of all problems in rigid body dynamics is the gyroscope. Much of the theory of the gyroscope has been covered in a four-volume treatise by F. Klein and A. Sommerfeld, *Theorie des Kreisels.*

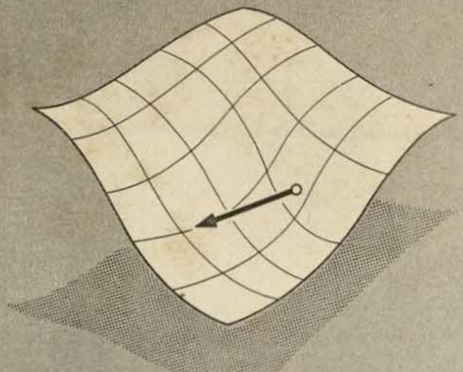
The main purpose of this modest chapter is to set up the exact equations of motion of a rigid body so that you can see just what is involved. We have all the equipment we need to do this. By a rigid body we mean a rigid assembly of particles, with fixed interparticle distances. We thereby exclude problems in which the body stretches or vibrates as a result of rotation. We shall solve several elementary problems, including a simple gyroscope and the problem of electron or nuclear spin resonance in a magnetic field.

---

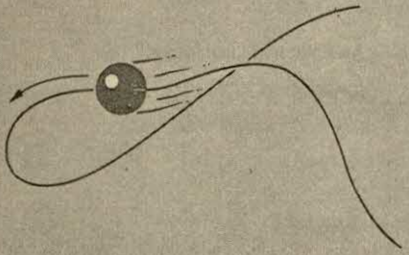
†On a first reading of this chapter we suggest that you stop at Eq. (6); on a second reading you may wish to stop at Eq. (47). The subsequent material is unusually difficult and demands much time and persistence.



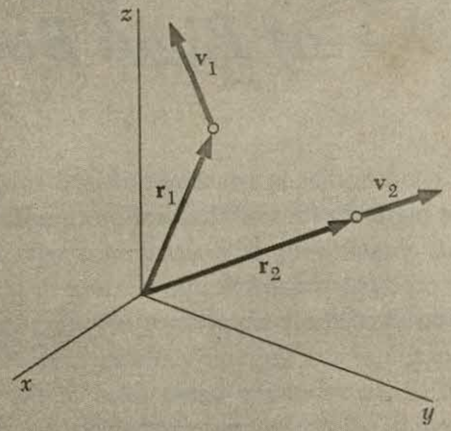
A particle moving with no constraints has three degrees of freedom ($f = 3$).



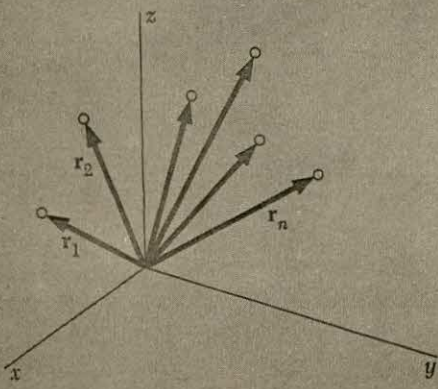
A particle constrained to move on a surface has only two degrees of freedom ($f = 2$).



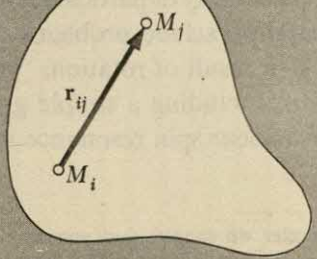
A bead constrained to slide on a wire has only one degree of freedom ($f = 1$).



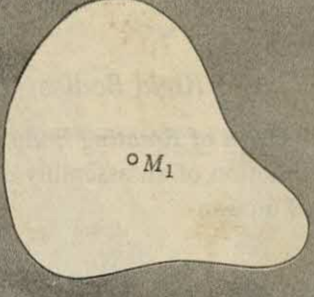
A system of two free particles has $f = (2 \times 3) =$ 6 degrees of freedom.



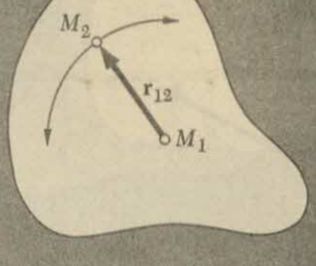
A system of N free particles has $f = 3N$ degrees of freedom.



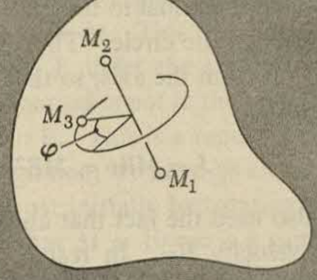
A rigid body has many particles, but there are also many constraints. The vector $\mathbf{r}_{ij}$ between two particles M_i and M_j is fixed. Thus $f \ll 3N$.



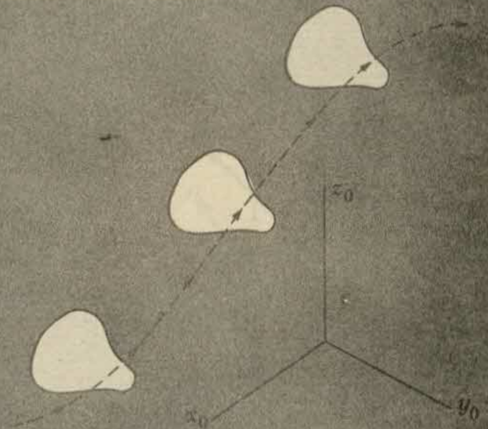
Suppose we choose one particle M_1 in the body. It has three degrees of freedom.



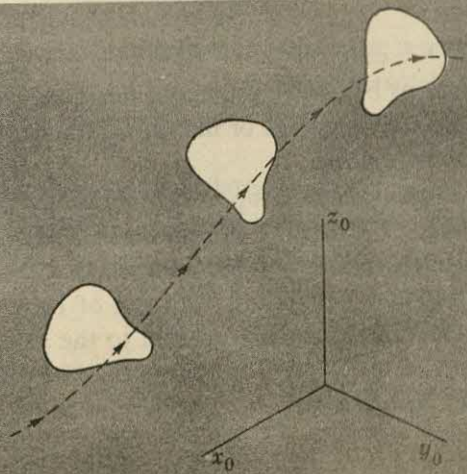
Once a frame is chosen in which M_1 is fixed, M_2 has only two degrees of freedom, since it must move on the surface of a sphere of radius $\mathbf{r}_{12}$.



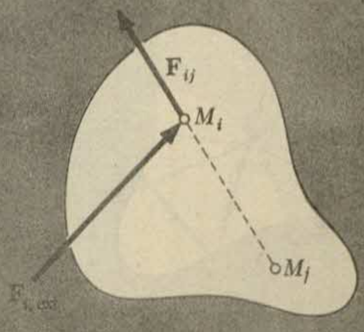
Once a frame is chosen in which M_1 and M_2 are fixed, any other particle M_3 in the body has only one degree of freedom, since it must move on the circle shown. Thus $f = 6$ for a rigid body.



Another view of this: Our rigid body can translate without rotating relative to inertial frame $(x_0y_0z_0)$. There are three degrees of translational freedom.



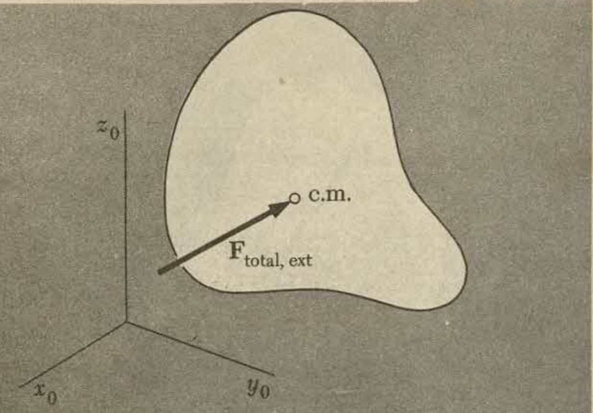
In addition it can rotate. There are three degrees of rotational freedom, so $f = 3 + 3 = 6$.



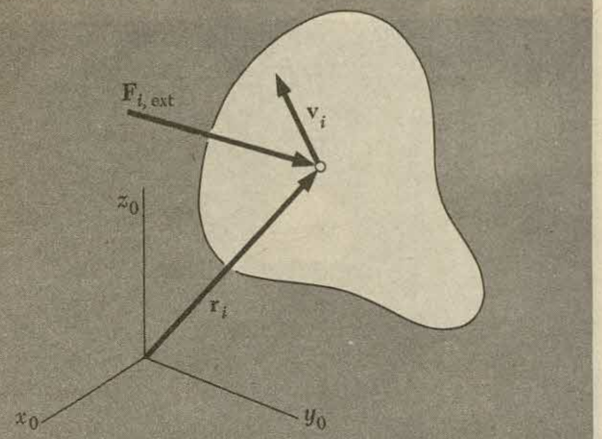
The force on each mass particle M_i is

$$\mathbf{F}_i = \mathbf{F}_{i,\mathrm{ext}} + \sum_{j \neq i} \mathbf{F}_{ij}.$$

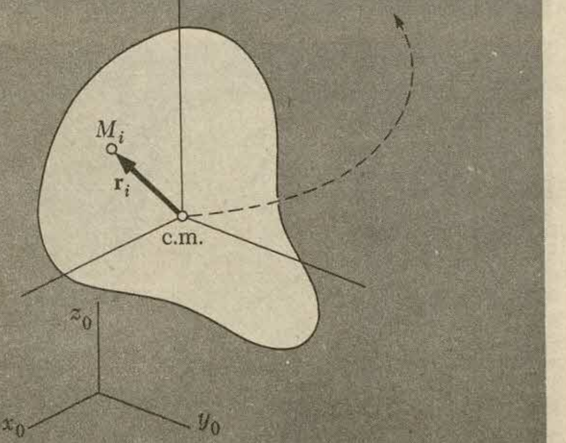
But if the internal forces obey Newton's third law, they cancel in pairs. Thus



$\mathbf{F}_{\text{ext}} = M\mathbf{a}_{\text{c.m.}}$. The center of mass translates as if the total external force were applied to *it,* and as if all of the mass were concentrated at the center of mass.



How does the body rotate? *Relative to the origin of the inertial frame* $d\mathbf{J}_{\text{tot}}/dt = \mathbf{N}_{\text{tot,ext}}$ if the internal forces are central and obey the third law.



If the body is translating, the origin of frame $x_0y_0z_0$ is very inconvenient. *Relative to the center of mass as origin,* we *also* have $d\mathbf{J}/dt = \mathbf{N}_{\text{tot,ext}}$! The center of mass is often a very convenient origin for discussion of rotations.

Equations of Motion of Rotating Body. We saw in Chap. 6 that the rotational motion of an assembly of particles is described by the equation of motion

$$\frac{d\mathbf{J}}{dt} = \mathbf{N}. \tag{1}$$

Here **J** is the angular momentum and **N** is the torque. Both **J** and **N** must be referred to a suitable common origin, which usually (but not always) will be the center of mass. Thus

$$\mathbf{J} = \Sigma M_n \mathbf{r}_n \times \mathbf{v}_n; \qquad \mathbf{N} = \Sigma \mathbf{r}_n \times \mathbf{F}_n; \tag{2}$$

here $\mathbf{F}_n$ is the external force on the nth particle.

The central physics of the chapter can be illustrated by a simple example, that of a thin circular hoop (of radius R) that rotates about an axis which is normal to the plane of the hoop and passes through the center of the circle. Then all of the mass M of the hoop is equidistant from the axis, so that the angular momentum **J** has the magnitude

$$J = MRv = MR^2\omega. \tag{3}$$

Here we have also used the fact that all points on the hoop move with the same velocity $R\omega$. In result (3) we see that MR^2 is a property of the hoop. It is usual to call

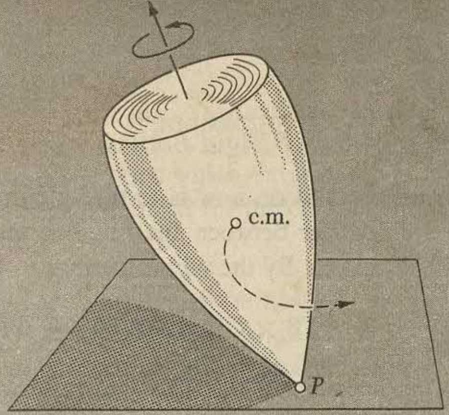
$$I = MR^2 \tag{4}$$

the *moment of inertia* of the hoop for rotation about the specified axis. Then $J = I\omega$ or (if we maintain the special requirement that the axis of rotation remain normal to the plane of the hoop) we may write

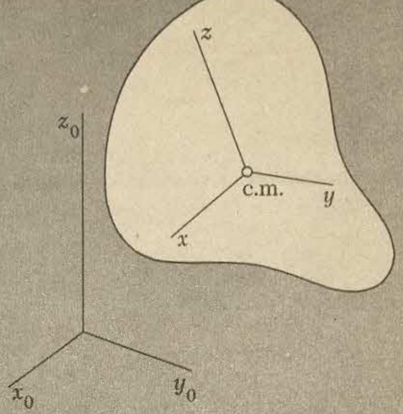
$$\mathbf{J} = I\boldsymbol{\omega}. \tag{5}$$

In the sections below we shall consider the generalization of the definition of the moment of inertia to bodies of arbitrary shape and mass distribution.

If the radius of the hoop is 100 cm and the mass 1 kg, then $I = (10^3 \text{ gm})(10^2 \text{ cm})^2 = 10^7 \text{ gm-cm}^2$. The angular momentum at an angular velocity of 100 rad/sec is $J = (10^7)(10^2) = 10^9$ gm-cm²/sec $= 10^9$ erg-sec. If a force of 10^4 dynes is applied at the rim (and in a direction parallel to the axis of rotation), the torque is $N = (10^2 \text{ cm})(10^4 \text{ dynes}) = 10^6$ dyne-cm. The direction of the torque is that of $\mathbf{r} \times \mathbf{F}$ and thus is normal to the axis of rotation.



If one point P is at rest (even if the center of mass is moving, as in the case of this top), then P (not the center of mass) is the appropriate origin of coordinates.



If we choose the center of mass as origin, we then construct a reference frame (xyz) fixed in the body. Reference frame $(x_0y_0z_0)$ is inertial and is not to be confused with (xyz).

From (1) we see that if the torque is applied only for a short time interval Δt, then $\Delta \mathbf{J} = \mathbf{N}\,\Delta t$. This tells us that the direction of $\mathbf{J}$ tilts to one side by $\Delta\mathbf{J}$ under the brief application of the torque; notice, however, that $\Delta\mathbf{J}$ is not in the direction of the force $\mathbf{F}$, but is at right angles to $\mathbf{F}$. This is a remarkable property of a rotating system. The argument we have given makes sense only if $\Delta J \ll J$, so the body must initially be rotating. For the numbers given above we see that $\Delta t = 10$ sec is a suitable short time interval, for then $N\,\Delta t = (10^6 \text{ dyne-cm})(10 \text{ sec}) = 10^7$ erg-sec, which is small in comparison with $J = 10^9$ erg-sec.

It is very important to be clear about the selection of the origin and the inertial frame. If the sum of the forces $(\Sigma \mathbf{F}_n)$ is zero, the proper origin will be the position of the center of mass

$$\mathbf{R}_{\text{c.m.}} = \frac{\Sigma M_n \mathbf{r}_n}{\Sigma M_n}, \tag{6}$$

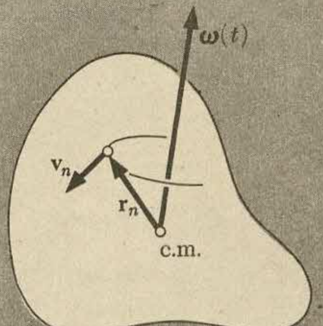
because an inertial frame always exists in which the center of mass of an assembly of particles is at rest, provided the total force vanishes. In general the body will rotate about an axis through this origin.

Let ω be the instantaneous angular velocity of the body about an axis through the origin. The velocity of the nth particle referred to the origin was expressed in Chap. 3 as

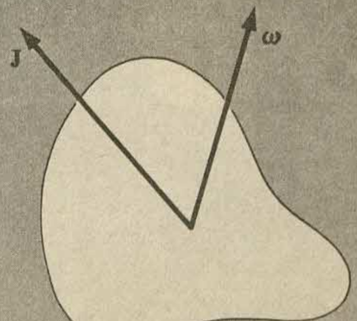
$$\mathbf{v}_n = \boldsymbol{\omega} \times \mathbf{r}_n, \tag{7}$$

so that (2) becomes

$$\mathbf{J} = \Sigma M_n \mathbf{r}_n \times \mathbf{v}_n = \Sigma M_n \mathbf{r}_n \times (\boldsymbol{\omega} \times \mathbf{r}_n). \tag{8}$$



At any instant the particles of the body may have angular velocity ω about the center of mass. Here we use body axes.



Since $\mathbf{J} = \omega\,\Sigma M_n r_n^2 - \Sigma M_n(\mathbf{r}_n \cdot \omega)\mathbf{r}_n$ relative to the body axes, $\mathbf{J}$ and ω need not be parallel.

What we want to do now is to find a compact way of writing this vector relation between the angular momentum **J** and the angular velocity $\boldsymbol{\omega}$. By the vector identity

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B}); \tag{9}$$

with the replacements $\mathbf{A} \rightarrow \mathbf{r}_n$; $\mathbf{B} \rightarrow \boldsymbol{\omega}$; $\mathbf{C} \rightarrow \mathbf{r}_n$, we have

$$\mathbf{r}_n \times (\boldsymbol{\omega} \times \mathbf{r}_n) = \boldsymbol{\omega} r_n^2 - \mathbf{r}_n(\mathbf{r}_n \cdot \boldsymbol{\omega}). \tag{10}$$

We substitute (10) in (8) to obtain

$$\mathbf{J} = \Sigma M_n[\boldsymbol{\omega} r_n^2 - \mathbf{r}_n(\mathbf{r}_n \cdot \boldsymbol{\omega})]. \tag{11}$$

All particles have the same $\boldsymbol{\omega}$, so that the result (11) may be written in terms of the three components of $\boldsymbol{\omega}$. Each component will be multiplied by a certain sum over the masses and positions of the particles. For example, the x component of (11) is

$$J_x = \omega_x \Sigma M_n r_n^2 - \Sigma x_n(\mathbf{r}_n \cdot \boldsymbol{\omega}). \tag{12}$$

On writing

$$\mathbf{r}_n \cdot \boldsymbol{\omega} = x_n\omega_x + y_n\omega_y + z_n\omega_z,$$

we have

$$J_x = \omega_x \Sigma M_n r_n^2 - \omega_x \Sigma M_n x_n^2 - \omega_y \Sigma M_n x_n y_n - \omega_z \Sigma M_n x_n z_n. \tag{13}$$

There are similar equations for the y and z components of (11).

Inertial Coefficients. We see that the expression (13) for J_x involves the three quantities

$$\Sigma M_n(r_n^2 - x_n^2); \qquad -\Sigma M_n x_n y_n; \qquad -\Sigma M_n x_n z_n.$$

These quantities depend on the distribution of mass in the body and on the instantaneous orientation of the body relative to the x, y, z coordinate axes which are to be viewed for the present as fixed in an inertial reference frame. Thus they will depend on time. We call these quantities *inertial coefficients* or *moments of inertia:*

$$\begin{aligned} I_{xx} &\equiv \Sigma M_n(r_n^2 - x_n^2); \\ I_{xy} &\equiv -\Sigma M_n x_n y_n; \\ I_{xz} &\equiv -\Sigma M_n x_n z_n, \end{aligned} \tag{14}$$

and similarly for the quantities which enter into J_y and J_z. We may then write the three equations for the components of (11) in the form

$$\begin{aligned} J_x &= I_{xx}\omega_x + I_{xy}\omega_y + I_{xz}\omega_z; \\ J_y &= I_{yx}\omega_x + I_{yy}\omega_y + I_{yz}\omega_z; \\ J_z &= I_{zx}\omega_x + I_{zy}\omega_y + I_{zz}\omega_z. \end{aligned} \tag{15}$$

We see that for a body of arbitrary shape and mass distribution the angular momentum **J** is not simply given by a scalar times the angular velocity ω. That is, **J** is not in general in the direction of ω. This circumstance is the source of the complex behavior exhibited by rotating bodies. The simple problems in rigid-body dynamics occur for spheres, for which we shall see that **J** is always parallel to ω. In the absence of a torque, **J** is constant in space, and for a general body the vector ω will precess about **J**.

The terms I_{xx}, I_{yy}, I_{zz} are called diagonal terms of the array (or matrix)

$$\begin{pmatrix} I_{xx} & I_{yx} & I_{zx} \\ I_{xy} & I_{yy} & I_{zy} \\ I_{xz} & I_{yz} & I_{zz} \end{pmatrix};$$

the other terms are called off-diagonal terms. Note that a diagonal term (e.g., I_{xx}), being defined as $\Sigma M_n(r_n^2 - x_n^2) = \Sigma M_n(y_n^2 + z_n^2)$, is the sum of the products of each mass by the square of its distance from the x axis; hence the designation "moment of inertia about the x axis." If $\rho(\mathrm{r})$ is the density at r, we may rewrite the I's in integral form as

$$I_{xx} = \int\rho(\mathrm{r})(r^2 - x^2)\,dV; \qquad I_{xy} = -\int\rho(\mathrm{r})xy\,dV, \qquad \text{etc.} \tag{16}$$

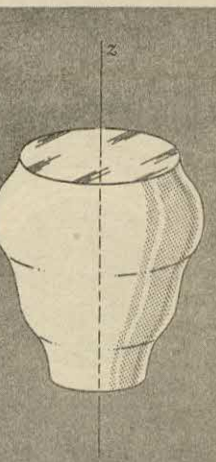
Here dV is the element of volume. We can check the change from the summation Σ over volume elements to the integral form by noting that $\int\rho(\mathrm{r})\,dV$ is equal to the mass contained in the volume element over which the integration is carried out.

We observe that the sum of the diagonal terms

$$I_{xx} + I_{yy} + I_{zz} = 2\Sigma M_n r_n^2 = 2\int\rho(\mathrm{r})r^2\,dV, \tag{17}$$

using (14) or (16) for I_{xx} and using the analogous expressions for I_{yy} and I_{zz}. Note that the sum in (17) is *isotropic*—that is, it is independent of the orientation of the body relative to the coordinate axes. Note also that the off-diagonal terms are symmetrical:

$$I_{xy} = I_{yx}; \qquad I_{xz} = I_{zx}; \qquad I_{yz} = I_{zy}. \tag{18}$$



Suppose we have a body with axial symmetry (e.g., about the z axis). Then $I_{xy} = 0$, because for any M_n at x_n, y_n there is a corresponding M_n at $-x_n$, y_n.

***Example. Showing* J *not parallel to* ω.** Two masses of 200 gm and 300 gm are separated by a light rod 50 cm in length. The center of mass of the system serves as the origin of a Cartesian coördinate system. The rod lies in the xy plane and makes an angle of 20° with the y axis. Find the inertial coefficients I_{xx} and I_{xy}.

Calculate first the distance of the center of mass from the 200-gm mass. We have, using (6),

$$R_{\text{c.m.}} = \frac{(200)(0) + (300)(50)}{200 + 300} = 30 \text{ cm.}$$

The Cartesian coordinates of the 200-gm mass (denoted as M_1) are referred to the center of mass as origin,

$$\begin{aligned} x_1 &= 30 \sin 20^\circ = 30(0.342) \cong 10.3 \text{ cm;} \\ y_1 &= 30 \cos 20^\circ = 30(0.940) \cong 28.2 \text{ cm;} \\ z_1 &= 0. \end{aligned}$$

The Cartesian coordinates of the 300-gm mass (denoted as M_2) are

$$\begin{aligned} x_2 &= -20 \sin 20^\circ \cong -6.8 \text{ cm;} \\ y_2 &= -20 \cos 20^\circ \cong -18.8 \text{ cm;} \\ z_2 &= 0. \end{aligned}$$

Using these values of the coordinates, we proceed to evaluate the inertial coefficients defined by (14):

$$\begin{aligned} I_{xx} &= M_1(r_1{}^2 - x_1{}^2) + M_2(r_2{}^2 - x_2{}^2) \\ &= 200[(30)^2 - (10.3)^2] + 300[(20)^2 - (6.8)^2] \\ &= 2.65 \times 10^5 \text{ gm-cm}^2; \\ I_{xy} &= -M_1x_1y_1 - M_2x_2y_2 \\ &= -(200)(28.2)(10.3) - (300)(6.8)(18.8) \\ &= -0.96 \times 10^5 \text{ gm-cm}^2. \end{aligned}$$

Now suppose that the rod rotates about the x axis with angular velocity ω. Find the components of **J**.

From (15) we have

$$J_x = I_{xx}\omega; \qquad J_y = I_{xy}\omega; \qquad J_z = 0,$$

where we have used the fact that $I_{zx} = 0$ for the rod in the assumed position because z_1 and z_2 are both zero. Then

$$\frac{J_y}{J_x} = \frac{I_{xy}}{I_{xx}} = \frac{(-0.96 \times 10^5)}{(2.65 \times 10^5)} = -0.363.$$

Thus we see that at the instant the rod lies in the xy plane the angular momentum vector also lies in the xy plane at an angle $\varphi = \tan^{-1}(-0.363) \cong -20°$ with the x axis. Thus $\mathbf{J}$ rotates about the x axis and is not parallel to ω, but is perpendicular to the rod.

How can we understand this result? Observe that the rod has zero moment of inertia (on our assumption of point masses) about its own length. Thus a component of ω parallel to the rod has no corresponding component of $\mathbf{J}$.

Example. Moment of inertia of spherical shell. Evaluate the inertial coefficients for a thin uniform spherical shell of mass density σ per unit area, for an axis through the center of the sphere.

Observe first that because of symmetry the off-diagonal coefficients I_{xy}, I_{yx}, I_{xz}, I_{zx}, I_{yz}, I_{zy} in (16) are all zero for a spherical distribution about the origin. This follows because for every contribution xy to the sum or integral defining I_{xy} there is for a sphere an equal but opposite contribution $(-x)y$. That is, the average $\langle xy\rangle$ over a spherical surface is zero.

Observe that $r^2 = x^2 + y^2 + z^2$, whence the average values over a spherical shell satisfy

$$\langle x^2\rangle = \tfrac{1}{3}\langle r^2\rangle. \tag{19}$$

Thus the diagonal coefficients I_{xx}, I_{yy}, I_{zz} are all equal, and because their sum is $2\Sigma M_n r_n{}^2$, they must each equal $\tfrac{2}{3}\Sigma M_n r_n{}^2$:

$$I = I_{xx} = I_{yy} = I_{zz} = \tfrac{2}{3}\Sigma M_n r_n{}^2 = \tfrac{2}{3}r^2\Sigma M_n = \tfrac{2}{3}4\pi\sigma r^4. \tag{20}$$

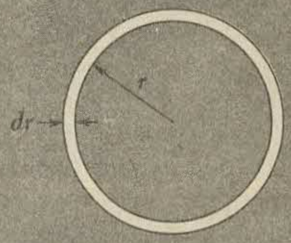
Note that the mass of the shell is $4\pi\sigma r^2$, with σ as the mass per unit area of the shell.

Example. Moment of inertia of sphere. Evaluate I for a sphere of uniform density ρ, for an axis through the center of the sphere.

We may assemble the sphere from spherical shells of thickness dr and having mass density $\sigma = \rho\, dr$ per unit area. Then from the result (20) we have for the sphere

$$I \equiv I_{xx} = I_{yy} = I_{zz} = \frac{2}{3}\int_0^R (4\pi r^4)(\rho\, dr)$$
$$= \frac{8\pi\rho}{3}\int_0^R r^4\, dr = \frac{8\pi}{15}\rho R^5 = \frac{2}{5}MR^2, \tag{21}$$

where $M = \dfrac{4\pi}{3}\rho R^3$ is the mass and R is the radius of the sphere.



$$M_{\text{shell}} = 4\pi\sigma r^2$$

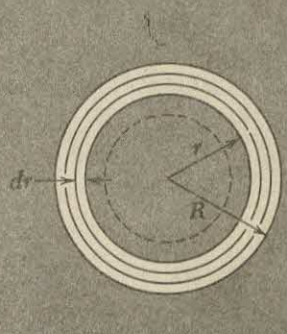
A uniform spherical shell is axially symmetric about x, y, and z, so

$$I_{xy} = I_{xz} = I_{yz} = 0.$$

Also by symmetry

$$I = I_{xx} = I_{yy} = I_{zz} = \tfrac{2}{3}r^2 M_{\text{shell}}.$$

Thus $I = (8\pi/3)\sigma r^4\, dr$; where σ is the mass per unit area of the shell.



$$M = \frac{4\pi}{3}\rho R^3$$

To find I for a uniform solid sphere we divide the sphere into concentric shells, and integrate:

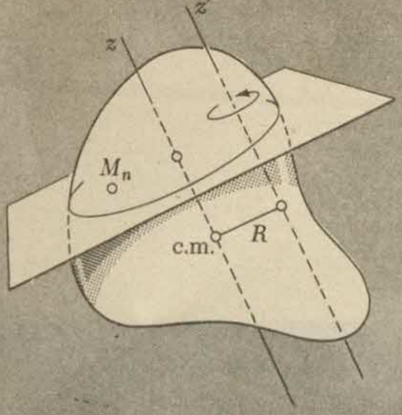
$$I = \frac{8\pi}{3}\rho\int_0^R r^4\, dr = \frac{8\pi}{15}\rho R^5 = \frac{2}{5}MR^2.$$

We observe that for a sphere or a spherical shell the angular momentum components given by (15) reduce to $\mathbf{J} = I\boldsymbol{\omega}$. Thus for these geometries the angular momentum is always parallel to the angular velocity.

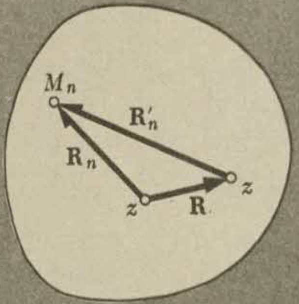
Example. Parallel-axis theorem. Evaluate I_{xx} for a general body about an axis parallel to $\hat{\mathbf{x}}$ and at a perpendicular distance a along the y axis from the center of mass.

In (14) we replace all y_n by $y_n + a$, leaving x_n and z_n unaltered. Then

$$\Sigma M_n r_n{}^2 \rightarrow \Sigma M_n (r_n{}^2 + 2y_n a + a^2), \tag{22}$$

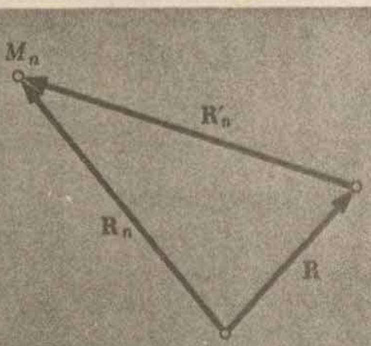


Suppose we wish to compute $I_{z'z'}$ where axis z' is parallel to z. How can we do this?



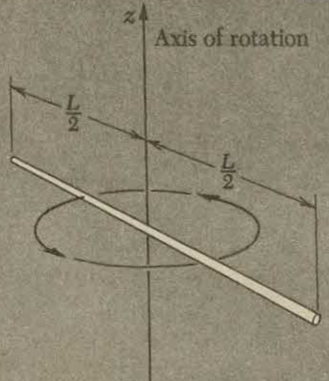
For any mass point M_n we can construct the plane normal to the zz' axes and passing through M_n (see preceding figure).

Then $R_n{}^2 = x_n{}^2 + y_n{}^2$. Thus . . .

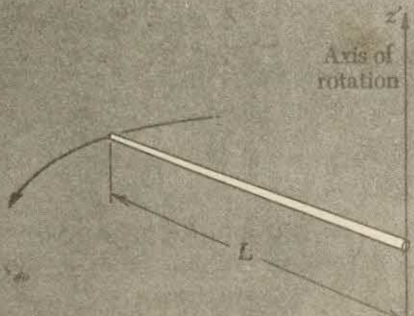


$$\begin{aligned} I_{z'z'} &= \Sigma M_n R'_n{}^2 = \Sigma M_n (\mathbf{R}_n - \mathbf{R})^2 \\ &= \Sigma M_n R_n{}^2 + R^2 \Sigma M_n - 2\Sigma M_n (\mathbf{R}_n \cdot \mathbf{R}). \end{aligned}$$

Since $\Sigma M_n \mathbf{R}_n = 0$, we have $I_{z'z'} = MR^2 + I_{zz}$; the parallel-axis theorem. Here $M = \Sigma M_n$ is the total mass.



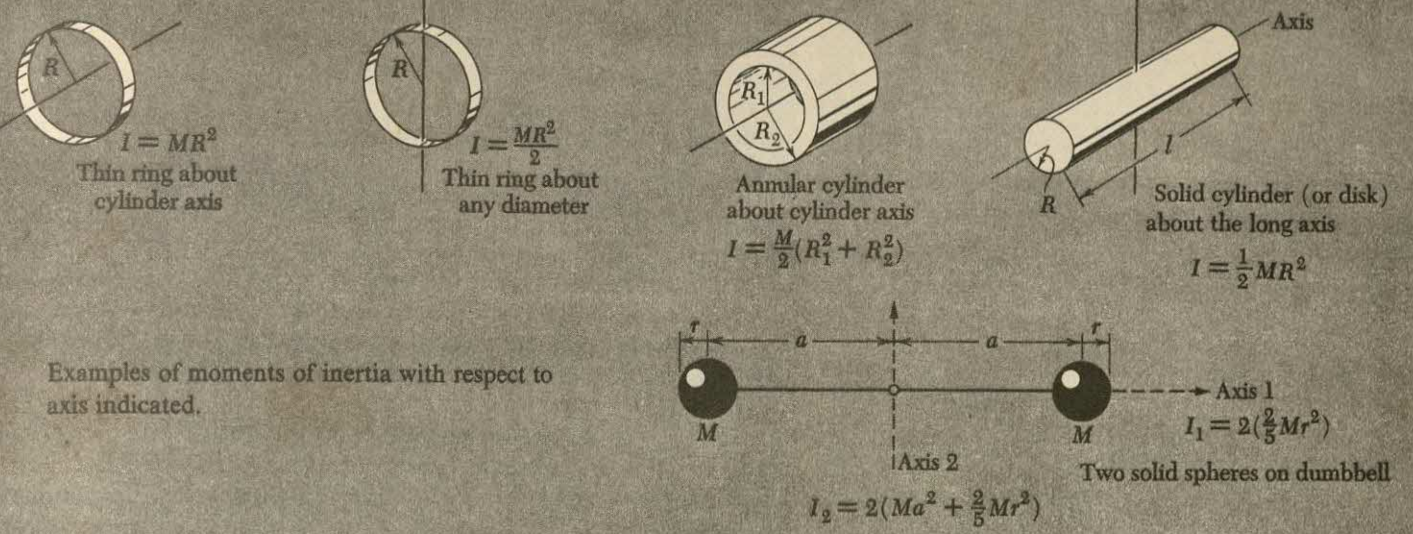
A thin rigid rod: $I_{zz} = 2\int_0^{L/2} x^2 \left(\frac{M}{L}\right) dx$, where M is total mass of the rod. Thus $I_{zz} = ML^2/12$.



Here $\quad I_{z'z'} = \int_0^L x^2 \left(\frac{M}{L}\right) dx = \frac{ML^2}{3};$

but $\quad \frac{ML^2}{3} = \frac{ML^2}{12} + M\left(\frac{L}{2}\right)^2,$

which verifies the parallel-axis theorem.



$I = MR^2$
Thin ring about cylinder axis

$I = \frac{MR^2}{2}$
Thin ring about any diameter

Annular cylinder about cylinder axis
$I = \frac{M}{2}(R_1^2 + R_2^2)$

Solid cylinder (or disk) about the long axis
$I = \frac{1}{2}MR^2$

$I_1 = 2(\frac{2}{5}Mr^2)$
Two solid spheres on dumbbell

$I_2 = 2(Ma^2 + \frac{2}{5}Mr^2)$

Examples of moments of inertia with respect to axis indicated.

but $\Sigma M_n y_n = 0$ by the choice of the center of mass. Thus the new value of I_{xx} is

$$I_{xx} = \Sigma M_n(r_n^2 + a^2 - x_n^2) = I^0_{xx} + a^2\Sigma M_n, \qquad (23)$$

where $I^0_{xx} = \Sigma M_n(r_n^2 - x_n^2)$ is the moment of inertia referred to an axis through the center of mass. We see that the moment of inertia about a given axis is equal to the moment of inertia about the parallel axis through the center of mass plus the moment of inertia about the given axis calculated as if the total mass of the body were located at the center of mass. This result is useful in treating rolling bodies. Because MR^2 is the moment of inertia of a hoop about an axis through its center and perpendicular to its plane, $2MR^2$ is the moment of inertia about a parallel axis on the rim of the hoop.

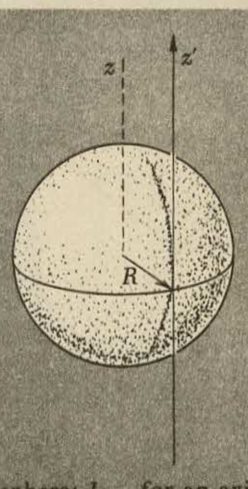
We see from (21) and (23) that the moment of inertia of a sphere of uniform density about an axis tangent to the surface is

$$I_{xx} = \tfrac{2}{5}MR^2 + MR^2 = \tfrac{7}{5}MR^2. \qquad (24)$$

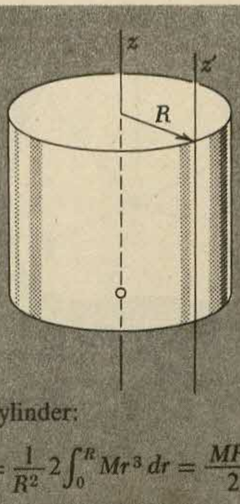
Summation Convention. The relations (14), (15), and (16) may be written in a very neat and compact form by using the summation convention† introduced in Mathematical Note 2, Chap. 2. We repeat the rules of the game:

1. Any Greek indices which occur twice in a product are to be summed over x, y, z. Thus

$$I_{\mu\nu}\omega_\nu \equiv I_{\mu x}\omega_x + I_{\mu y}\omega_y + I_{\mu z}\omega_z. \qquad (25)$$



A uniform sphere: $I_{z'z'}$ for an axis tangent to the surface is $I_{z'z'} = \frac{2}{5}MR^2 + MR^2 = \frac{7}{5}MR^2$.



A uniform cylinder:

$$I_{zz} = \frac{1}{R^2}2\int_0^R Mr^3\,dr = \frac{MR^2}{2};$$

$$I_{z'z'} = MR^2 + \frac{MR^2}{2} = \frac{3}{2}MR^2.$$

---

† Sometimes called the tensor summation convention.

2. An index which does not occur twice in a product can stand for x or y or z. Thus the equation

$$J_\mu = I_{\mu\nu}\omega_\nu \tag{26}$$

is equivalent to the entire set of equations (15). First, for example, we can let μ stand for x; and then (26) becomes

$$J_x = I_{x\nu}\omega_\nu \equiv I_{xx}\omega_x + I_{xy}\omega_y + I_{xz}\omega_z. \tag{27}$$

It is much neater to write (26) than to write the three equations (15).

3. By x_μ we mean either x or y or z. Now in the product $x_\mu x_\mu$ the index μ occurs twice, so that

$$x_\mu x_\mu \equiv x^2 + y^2 + z^2 = r^2. \tag{28}$$

But in $x_\mu x_\nu$ no Greek index occurs twice, and so $x_\mu x_\nu$ might stand for xx or xy or yz, etc. With the summation convention, (16) assumes the form

$$I_{\mu\nu} = \int \rho(\mathbf{r})\,(x_\alpha x_\alpha \delta_{\mu\nu} - x_\mu x_\nu)\,dV. \tag{29}$$

The $\delta_{\mu\nu}$ is the Kronecker delta symbol; it is zero unless $\mu = \nu$, when it has the value one. Suppose we are concerned with I_{xx}; we read the $x_\alpha x_\alpha \delta_{\mu\nu}$ term in (29) as r^2, and we read the $x_\mu x_\nu$ term as x^2. You can get along perfectly well without the summation convention. You should use it to save writing *only* if you are sure of what you are doing.

Kinetic Energy of Rotation. The kinetic energy of a rigid body referred to the center of mass at rest is called the rotational energy of the body. It is given by

$$K = \tfrac{1}{2}\Sigma M_n v_n{}^2 = \tfrac{1}{2}\Sigma M_n(\boldsymbol{\omega} \times \mathbf{r}_n)^2 = \tfrac{1}{2}\int \rho(\mathbf{r})\,(\boldsymbol{\omega} \times \mathbf{r})^2\,dV, \tag{30}$$

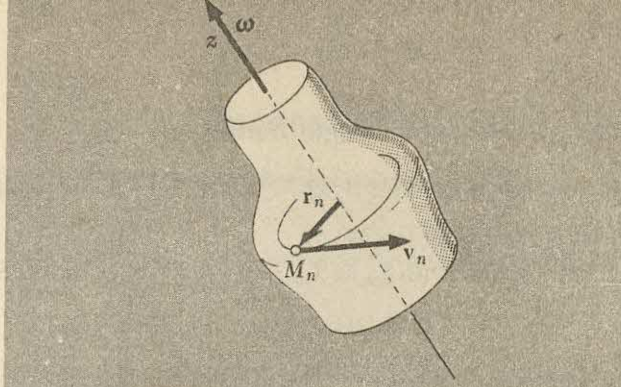
using $\mathbf{v}_n = \boldsymbol{\omega} \times \mathbf{r}_n$. Here $(\boldsymbol{\omega} \times \mathbf{r}_n)^2 \equiv (\boldsymbol{\omega} \times \mathbf{r}_n) \cdot (\boldsymbol{\omega} \times \mathbf{r}_n)$.

Consider first the rotational kinetic energy of a uniform sphere. Let ω be parallel to the z axis. Then

$$(\boldsymbol{\omega} \times \mathbf{r})^2 = \omega^2(x^2 + y^2), \tag{31}$$

but the moment of inertia of a sphere of uniform density ρ is given from (16) by

$$I = I_{zz} = \rho\int(x^2 + y^2)\,dV, \tag{32}$$



The kinetic energy of an axially symmetric body rotating about the axis of symmetry is

$$K = \tfrac{1}{2}I_{zz}\omega^2.$$

so that (30) reduces to

$$K = \tfrac{1}{2}\rho\int(\omega \times \mathbf{r})^2\,dV = \tfrac{1}{2}\omega^2\rho\int(x^2 + y^2)\,dV = \tfrac{1}{2}I\omega^2. \quad (33)$$

We now proceed to obtain an expression for the rotational kinetic energy of a rigid body of arbitrary shape. The result will be

$$K = \tfrac{1}{2}(\omega_x^2 I_{xx} + \omega_y^2 I_{yy} + \omega_z^2 I_{zz} + 2\omega_x\omega_y I_{xy} + 2\omega_y\omega_z I_{yz} + 2\omega_z\omega_x I_{zx}). \quad (34)$$

Using the vector relation

$$(\mathbf{A} \times \mathbf{B}) \cdot (\mathbf{C} \times \mathbf{D}) = (\mathbf{A} \cdot \mathbf{C})(\mathbf{B} \cdot \mathbf{D}) - (\mathbf{A} \cdot \mathbf{D})(\mathbf{B} \cdot \mathbf{C}), \quad (35)$$

we have

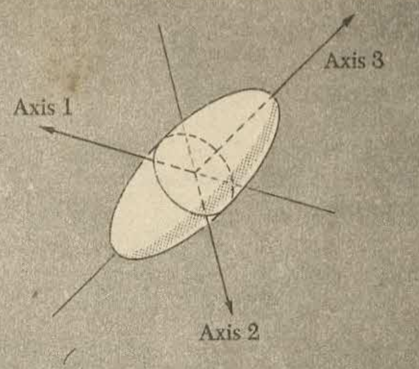
$$(\omega \times \mathbf{r})^2 = \omega^2 r^2 - (\omega \cdot \mathbf{r})^2. \quad (36)$$

According to the summation convention, (36) may be written as†

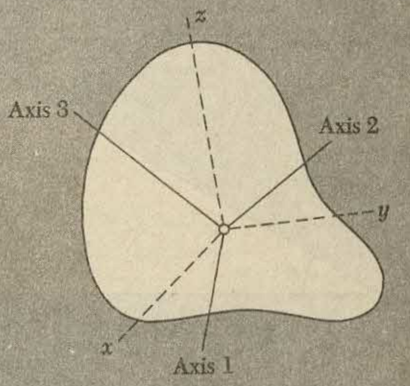
$$(\omega \times \mathbf{r})^2 = \omega_\mu\omega_\mu x_\nu x_\nu - \omega_\mu x_\mu \omega_\nu x_\nu = \omega_\mu\omega_\nu(x_\alpha x_\alpha \delta_{\mu\nu} - x_\mu x_\nu). \quad (37)$$

[This relation could easily have been written down directly, without using (36), by making use of the symbol $\epsilon_{\lambda\mu\nu}$, defined in Mathematical Note 2, Chap. 2.] When matters reach about the present degree of complication, the summation convention tends to become more convenient than vector notation.

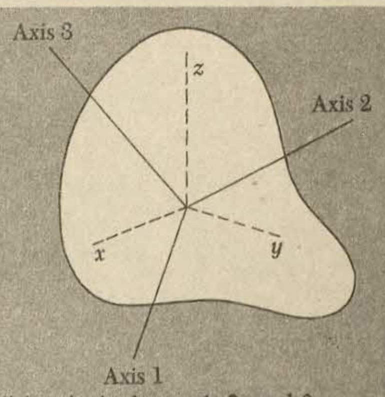
On substituting (37) into (30) and using (29), we have for the kinetic energy



For the *general* rigid body $K = \frac{1}{2}I_{\mu\nu}\omega_\mu\omega_\nu$. The surface $\frac{1}{2}I_{\mu\nu}\omega_\mu\omega_\nu = \text{const}$ is an ellipsoid in $\omega_x\omega_y\omega_z$ space. The principal axes (1,2,3) of the ellipsoid are shown.



If we make a suitable rotation of body axes (xyz), they will coincide with the principal axes of the ellipsoid. From now on we use principal body axes.



With principal axes 1, 2, and 3,

$$K = \tfrac{1}{2}I_1\omega_1^2 + \tfrac{1}{2}I_2\omega_2^2 + \tfrac{1}{2}I_3\omega_3^2$$

and

$$\mathbf{J} = I_1\omega_1 + I_2\omega_2 + I_3\omega_3.$$

All cross terms vanish!

---

† In case you have any doubt about how to read (37), we write it with the summations explicitly indicated:

$$(\omega \times \mathbf{r})^2 = \Big(\sum_\mu \omega_\mu^2\Big)\Big(\sum_\nu x_\nu^2\Big) - \Big(\sum_\mu \omega_\mu x_\mu\Big)\Big(\sum_\nu \omega_\nu x_\nu\Big),$$

where μ, ν run over x, y, z. Note that $\Sigma x_\nu^2 = r^2$.

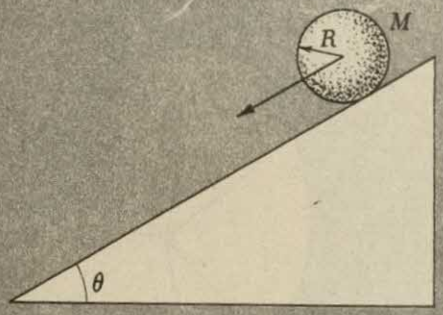
$$K = \tfrac{1}{2}\omega_\mu\omega_\nu \int \rho(\mathbf{r})\,(x_\alpha x_\alpha \delta_{\mu\nu} - x_\mu x_\nu)\,dV = \tfrac{1}{2} I_{\mu\nu}\omega_\mu\omega_\nu. \tag{38}$$

This is equivalent to (34).[†]

For a sphere with $I_{xx} = I_{yy} = I_{zz} = I$ we see that (38) reduces to

$$K = \tfrac{1}{2} I(\omega_x^2 + \omega_y^2 + \omega_z^2) = \tfrac{1}{2} I\omega^2, \tag{39}$$

in agreement with (33). The inertial coefficients are useful in the description both of the angular momentum and of the rotational energy of a rigid body.



Sphere rolling down inclined plane.

Example. Rolling sphere. A sphere of radius R and mass M rolls without slipping down a plane, of length L, which is inclined at an angle θ from the horizontal. What is the velocity of the center of mass of the sphere at the bottom end of the plane?

This is a standard type of textbook problem. One way to solve it is to equate the potential energy at the top end of the plane

$$U = Mgh = MgL\sin\theta \tag{40}$$

to the kinetic energy at the bottom end, which is

$$K = \tfrac{1}{2}Mv^2 + \tfrac{1}{2}I\omega^2, \tag{41}$$

the sum of the translational kinetic energy and of the rotational kinetic energy. Note that the two terms occurring in (41) are completely independent. The first term gives the translational kinetic energy due to the motion of the center of mass; the second term gives the kinetic energy due to the rotation of the body *about* the center of mass. It is intuitively clear that these terms are independent, as it is possible for the sphere to slide down the plane on one point of contact without rotating, and it is equally possible for the sphere just to rotate with its center of mass remaining stationary. In the motion of the sphere down the plane any combination of these two independent motions is possible.

There is one particular combination in which we are most often interested. It is called rolling without slipping. You have been rolling things since your first days of consciousness. Rolling without slipping is defined as that motion characterized by the point of contact with the plane being instantaneously at rest. It may also be defined by the condition that the velocity of the center of mass equals $R\omega$. Thus (41) may be written as

---

[†] It is worth noting that for a rigid body $J_\nu = \partial K/\partial\omega_\nu$. This may be seen by comparing Eqs. (26) and (38), or Eqs. (15) and (34).

$$K = \frac{1}{2}Mv^2 + \frac{1}{2}I\left(\frac{v}{R}\right)^2 = \frac{1}{2}Mv^2 + \frac{1}{2}\cdot\frac{2}{5}Mv^2 = \frac{7}{10}Mv^2, \tag{42}$$

where we have used $I = \frac{2}{5}MR^2$ for a uniform sphere. On equating (40) to (42) we have

$$v^2 = \tfrac{10}{7}\, gL \sin\theta, \tag{43}$$

at the bottom end.

Another way to solve the same problem is to exploit the fact that a rolling body at any instant may be viewed as rotating about the point of contact. The point of contact on a rolling body is always at rest. From this viewpoint the total kinetic energy is rotational about the point of contact with $I = \frac{7}{5}MR^2$, as given by (24) for rotation about an axis tangent to the surface. Then

$$K = \tfrac{1}{2}\cdot\tfrac{7}{5}MR^2\omega^2 = \tfrac{7}{10}Mv^2, \tag{44}$$

in agreement with (42).

Principal Axes. Notice the complicated general expression (34) for the rotational kinetic energy. This expression, which is for an arbitrary set of Cartesian axes, may always be simplified to three terms. The trick (and for most regular bodies it is a simple one) is to choose the proper set of axes. In this preferred set of axes called the *principal axes*, only the three diagonal coefficients $I_{xx} \equiv I_1$; $I_{yy} \equiv I_2$; $I_{zz} \equiv I_3$ occur. The off-diagonal coefficients conveniently vanish, so that the rotational kinetic energy is

$$K = \tfrac{1}{2}I_1\omega_1^2 + \tfrac{1}{2}I_2\omega_2^2 + \tfrac{1}{2}I_3\omega_3^2. \tag{45}$$

(For a cylinder the principal axes would be the longitudinal axis of the cylinder and two mutually perpendicular axes drawn from the longitudinal axis and perpendicular to it.) The equations (15) connecting $\mathbf{J}$ and $\boldsymbol{\omega}$ reduce to

$$J_1 = I_1\omega_1; \qquad J_2 = I_2\omega_2; \qquad J_3 = I_3\omega_3, \tag{46}$$

so that the rotational kinetic energy (45) may be written as

$$K = \frac{1}{2I_1}J_1^2 + \frac{1}{2I_2}J_2^2 + \frac{1}{2I_3}J_3^2. \tag{47}$$

Notice that **J** is parallel to ω if the body is rotating about a principal axis.

Euler Equations. The equation of motion

$$\frac{d}{dt}\mathbf{J} = \mathbf{N} \tag{48}$$

holds in an inertial reference frame. The inertial coefficients $I_{\mu\nu}$ are most conveniently defined for coordinate axes at rest in the rotating body, which is a noninertial frame. (Up to this point in the development we have not had to specify whether the axes were at rest or not.) By the argument of Chap. 3 we know how to transform a vector from an inertial frame to a rotating frame:

$$\left(\frac{d\mathbf{J}}{dt}\right)_I = \frac{d\mathbf{J}}{dt} + \boldsymbol{\omega} \times \mathbf{J}, \tag{49}$$

where ω is the angular velocity of the rotating frame. All other quantities on the right-hand side are as observed from the rotating frame. Thus (48) becomes

$$\frac{d\mathbf{J}}{dt} + \boldsymbol{\omega} \times \mathbf{J} = \mathbf{N}, \tag{50}$$

for **J** observed from the rotating frame.

We suppose that the Cartesian axes in the rotating frame lie along the principal axes 1, 2, 3. The component of (50) along axis 1 is, using (46),

$$\begin{aligned}\left(\frac{dJ}{dt}\right)_1 + (\boldsymbol{\omega} \times \mathbf{J})_1 &= \frac{dJ_1}{dt} + \omega_2 J_3 - \omega_3 J_2 \\ &= I_1 \frac{d\omega_1}{dt} + \omega_2 I_3 \omega_3 - \omega_3 I_2 \omega_2 = N_1, \quad (51)\end{aligned}$$

or, rearranging (51) and writing out also the equations for the components of (50) along axes 2 and 3,

$$I_1 \frac{d\omega_1}{dt} + (I_3 - I_2)\omega_2\omega_3 = N_1; \tag{52a}$$

$$I_2 \frac{d\omega_2}{dt} + (I_1 - I_3)\omega_1\omega_3 = N_2; \tag{52b}$$

$$I_3 \frac{d\omega_3}{dt} + (I_2 - I_1)\omega_1\omega_2 = N_3. \tag{52c}$$

These equations are known as the Euler equations. They are a good starting point for rotating-body problems. Remember always in using the Euler equations that the principal axes 1, 2, 3 are fixed in the body.

Example. Precession of sphere, no torque acting. For a uniform sphere with $I_1 = I_2 = I_3$ the Euler equations reduce to

$$I\dot{\omega}_1 = N_1; \qquad I\dot{\omega}_2 = N_2; \qquad I\dot{\omega}_3 = N_3. \tag{53}$$

In free motion $\mathbf{N} = 0$, and (53) tells us that $\boldsymbol{\omega} =$ const. The result $\boldsymbol{\omega} =$ const is a special feature of the free rotating sphere.

Example. Symmetrical top. A symmetrical top has $I_1 = I_2 \neq I_3$. We see from (52) that if no external torques act, then ω_3 (and only ω_3) is constant. Equations (52*a*) and (52*b*) become

$$\dot{\omega}_1 + \Omega\omega_2 = 0; \qquad \dot{\omega}_2 - \Omega\omega_1 = 0, \tag{54}$$

where

$$\Omega \equiv \frac{I_3 - I_1}{I_1}\omega_3, \tag{55}$$

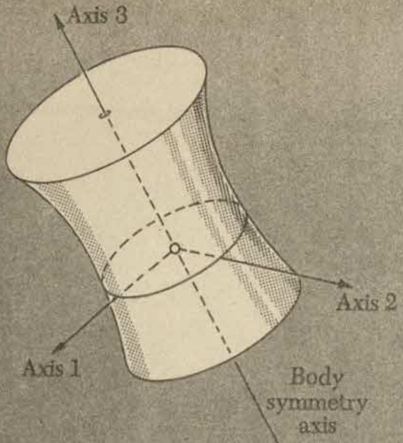
using $I_1 = I_2$.

A solution of (54) is given by

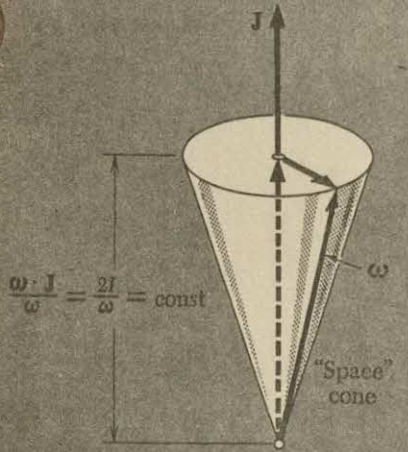
$$\omega_1 = A\cos\Omega t; \qquad \omega_2 = A\sin\Omega t, \tag{56}$$

where A is a constant. We see that the component of the angular velocity perpendicular to the figure axis (axis 3) of the top rotates with a constant angular velocity Ω. The component ω_3 of the angular velocity along the figure axis is constant. Therefore the vector $\boldsymbol{\omega}$ rotates uniformly with angular velocity Ω about the figure axis of the top. In other words, a top which spins about its figure axis with angular velocity ω_3 in force-free space will wobble with the frequency Ω given by (55).

For the earth I_3 is not exactly equal to I_1 because the earth is not exactly a sphere. The wobble (56) is actually very well observed, giving rise to what is called the variation of latitude. The wobble is so interesting that the International Latitude Service maintains a number of observatories just for the purpose of measuring it. One of them is at Ukiah, in Northern California. Equation (55) predicts for the earth a period of 305 days. The observed motion has an annual component (interpreted as



Consider the "free" rotation ($N = 0$) of a body with axial symmetry (e.g., z axis). Then $I_1 = I_2$, so $\omega_3 = \text{const}$.

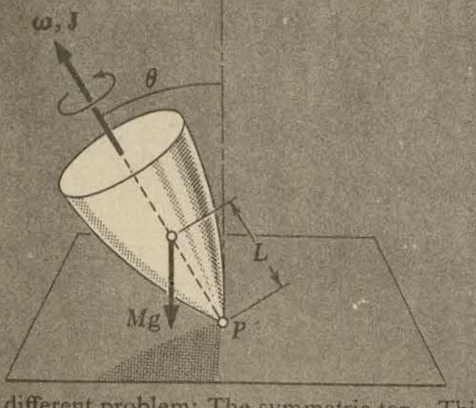

J
$\frac{\boldsymbol{\omega} \cdot \mathbf{J}}{\omega} = \frac{2I}{\omega} = \text{const}$
ω
"Space" cone

Also since $\mathbf{N} = 0$, $\mathbf{J}$ is fixed in space. The projection of ω on $\mathbf{J}$ is a constant

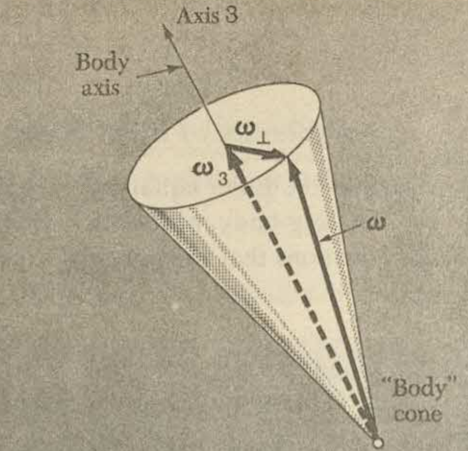
$[K = \frac{1}{2}I_1(\omega_1{}^2 + \omega_2{}^2) + \frac{1}{2}I_3\omega_3{}^2$, and

$\omega_1{}^2 + \omega_2{}^2 = \text{const}]$.

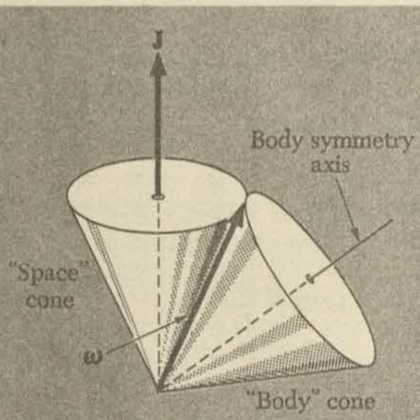
Thus $\boldsymbol{\omega}$ precesses at constant rate around $\mathbf{J}$.



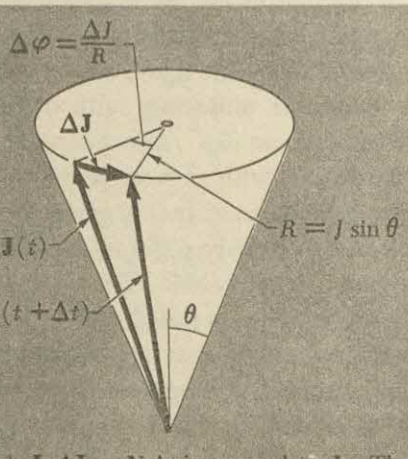
A different problem: The symmetric top. This time $|d\mathbf{J}/dt| = |\mathbf{N}| = MgL \sin\theta \neq 0$. If the top is spinning rapidly, ω and $\mathbf{J}$ approximately coincide with the figure axis.



In these circumstances $|\omega| = \text{const}$. The vector $\boldsymbol{\omega}$ precesses at a constant rate around the body axis of symmetry.



Thus the body cone rolls on the space cone (without slipping) for the case of "free" rotation of an axially symmetric body.



Since $\mathbf{N} \perp \mathbf{J}$, $\Delta\mathbf{J} = \mathbf{N}\,\Delta t$ is normal to $\mathbf{J}$. Thus $|J| = \text{const}$, but $\mathbf{J}$ precesses around the vertical axis. The precession frequency is

$$\Omega = \frac{\Delta\varphi}{\Delta t} = \frac{\Delta J}{\Delta t}\frac{1}{J \sin\theta} = \frac{MgL \sin\theta}{J \sin\theta} = \frac{MgL}{J}.$$

a forced oscillation with meteorological input) and a free period of 420 days. It was considered a triumph when Newcomb, at the end of the nineteenth century, accounted for the lengthening of the period from 305 to 420 days by the deformation of the earth under the effect of change in direction of centrifugal force. This provided one of the first estimates of the rigidity of the earth.

The whole subject is fascinating.† The free component of the motion appears to be damped, with a damping time of 30 years or less, which provides information on the nonelastic characteristics of the earth. However, no one has yet succeeded in explaining what maintains the motion.

Equation (50) or the Euler equations show us that the angular velocity components and hence the angular momentum components (referred to axes fixed on the body) are not constant, even when the torque is zero. But if there is no torque, the *magnitude* of the angular momentum is constant:

$$J_1^2 + J_2^2 + J_3^2 = \text{const.} \qquad (57)$$

To prove this result consider (50) with $\mathbf{N} = 0$:

$$\dot{\mathbf{J}} = -\boldsymbol{\omega} \times \mathbf{J}. \qquad (58)$$

Now with this result we have

$$\frac{d}{dt}(\mathbf{J} \cdot \mathbf{J}) = 2\mathbf{J} \cdot \dot{\mathbf{J}} = -2\mathbf{J} \cdot \boldsymbol{\omega} \times \mathbf{J} = 0, \qquad (59)$$

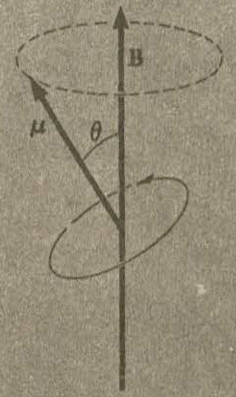
because $\boldsymbol{\omega} \times \mathbf{J}$ is perpendicular to $\mathbf{J}$. Thus

$$\mathbf{J} \cdot \mathbf{J} = \text{const} = J_1^2 + J_2^2 + J_3^2. \qquad Q.E.D. \qquad (60)$$

Spin Precession in a Constant Magnetic Field. By spin we mean the angular momentum of a particle about its center of mass. The precession of the angular momentum vector in a magnetic field is an interesting problem of great importance to atomic and solid state physics and to chemistry, biology, and geology. We make two assumptions:

1. The magnetic moment $\boldsymbol{\mu}$ of an elementary particle equals a constant times the angular momentum $\mathbf{J}$:

$$\boldsymbol{\mu} = \gamma \mathbf{J}. \qquad (61)$$



A similar problem: A magnetic moment $\boldsymbol{\mu} = \gamma\mathbf{J}$ in a magnetic field, where the torque $\mathbf{N} = \boldsymbol{\mu} \times \mathbf{B}$ is normal to $\mathbf{J}$. The precession frequency is $\omega = \gamma B$, again independent of θ.

---

† W. Munk and G. F. MacDonald, *The Rotation of the Earth* (Cambridge University Press, New York, 1961).

The constant γ is called the magnetomechanical ratio. For a free electron $\gamma \cong e/mc$.

2. The torque on the magnetic moment in the magnetic field $\mathbf{B}$ is equal to $\boldsymbol{\mu} \times \mathbf{B}$. This result is derived in Vol. II, but it is a matter of common experience that a magnetic field exerts a torque on a bar magnet or on a current loop.

Thus in the laboratory reference system we have, from $\dot{\mathbf{J}} = \mathbf{N}$, the equation

$$\boxed{\dot{\mathbf{J}} = \boldsymbol{\mu} \times \mathbf{B} = \gamma \mathbf{J} \times \mathbf{B}.} \tag{62}$$

Now suppose that $\mathbf{B} = B\hat{\mathbf{z}}$; then (62) becomes

$$\dot{J}_x = \gamma J_y B; \qquad \dot{J}_y = -\gamma J_x B; \qquad \dot{J}_z \doteq 0. \tag{63}$$

These equations are similar to (54). A solution of (63) is, with constant A,

$$J_x = A \sin \Omega t; \qquad J_y = A \cos \Omega t; \qquad J_z = \text{const}, \tag{64}$$

provided

$$\Omega = \gamma \mathrm{B}. \tag{65}$$

We call Ω the free precession frequency.

Spin Resonance. Now consider the motion of the spin in a magnetic field which has a constant z component denoted by B and a small alternating x component H_1 at frequency ω. Let the sum of these two fields be

$$\mathbf{B} = H_1 \sin \omega t\, \hat{\mathbf{x}} + \mathrm{B}\hat{\mathbf{z}}. \tag{66}$$

The three components of the equation of motion (62) are

$$\dot{J}_x = \Omega J_y; \qquad \dot{J}_y = -\Omega J_x + \gamma J_z H_1 \sin \omega t;$$
$$\dot{J}_z = -\gamma J_y H_1 \sin \omega t. \tag{67}$$

Although these equations can be solved exactly, it is convenient to make an approximation. We suppose that $J_y \ll J_z$, which corresponds to the spin making a small angle with the z axis. Then we may set $\dot{J}_z \cong 0$, so that $J_z = J$, a constant.

We now look for solutions for J_x, J_y which have the same time dependence as the driving field (66):

$$J_x = A \sin \omega t; \qquad J_y = C \cos \omega t, \tag{68}$$

where A and C are constants. The sine and cosine forms are a guess based on our picture of the motion of **J** as a rotation about the z axis. Substituting (68) into (67), we have

$$\begin{aligned} \omega A \cos \omega t &= \Omega C \cos \omega t - \omega C \sin \omega t \\ &= -\Omega A \sin \omega t + \gamma J H_1 \sin \omega t. \end{aligned} \tag{69}$$

It follows that

$$\omega A = \Omega C; \qquad -\omega C = -\Omega A + \gamma J H_1 \tag{70}$$

or

$$C = \frac{\gamma \omega J}{\omega^2 - \Omega^2} H_1. \tag{71}$$

Thus, from (68),

$$J_x = \frac{\gamma \Omega J}{\omega^2 - \Omega^2} H_1 \sin \omega t; \qquad J_y = \frac{\gamma \omega J}{\omega^2 - \Omega^2} H_1 \cos \omega t. \tag{72}$$

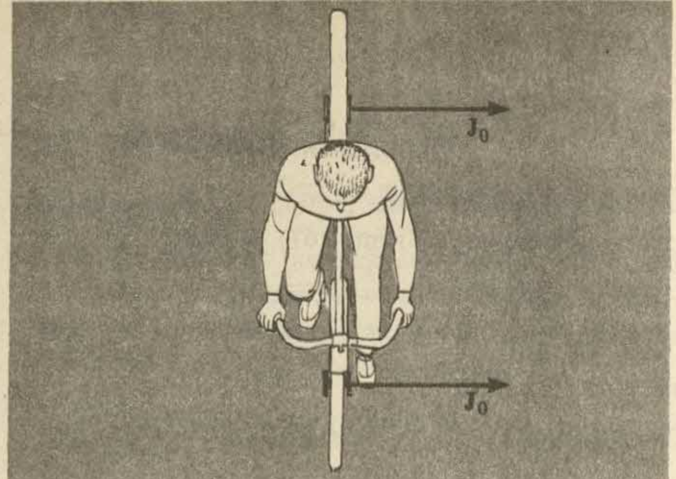
This result exhibits a resonance maximum when $\omega = \Omega \equiv \gamma B$. [The infinity or singularity here is removed in the exact solution of (67) in which $\dot{J}_z$ is not neglected.] We see that the system responds in both the x and y directions to an alternating magnetic field in the x direction.

The phenomena of electron and nuclear spin resonance are rich in applications. The obvious first application to nuclear physics is in the determination of the value of the magnetomechanical ratio $\gamma = \mu/J$ for a nucleus. To do this we measure the frequency and magnetic field strength at which resonance is observed, for these are related by

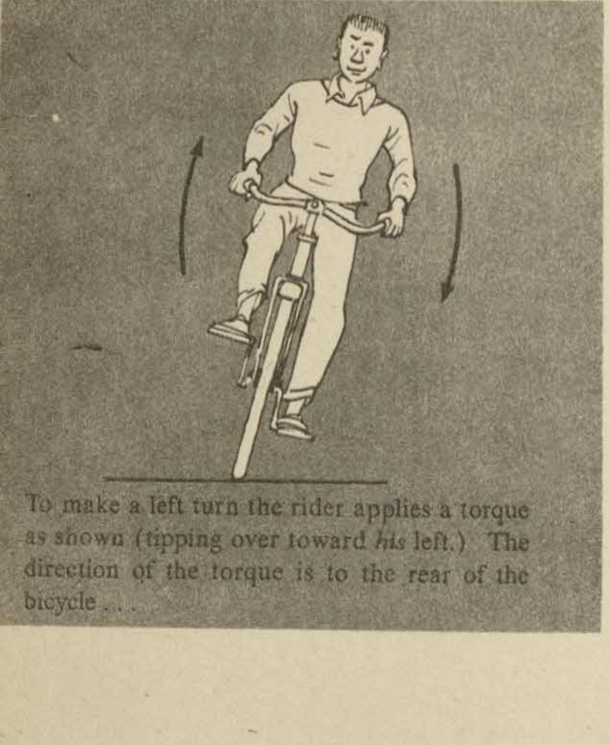
$$\omega = \gamma B. \tag{73}$$

Elementary Gyroscope. Consider the motion under gravity of a symmetrical top which (in our special case) spins with angular velocity ω_3 about a horizontal axis. The top is supported at one end of a frictionless pivot. The force of gravity exerts a torque of fixed magnitude N about the pivot. The direction of **N** is normal to the axis of the top and normal to the vertical. We assert the startling result that the top is stable when its axis is horizontal. We wish to find the precessional angular velocity Ω of the axis as it slowly sweeps around in the horizontal plane. We make the approximation that ω_3 is constant, independent of the precessional angular velocity. This is a good approximation in the limit in which the top spins very rapidly as compared with the precessional frequency.

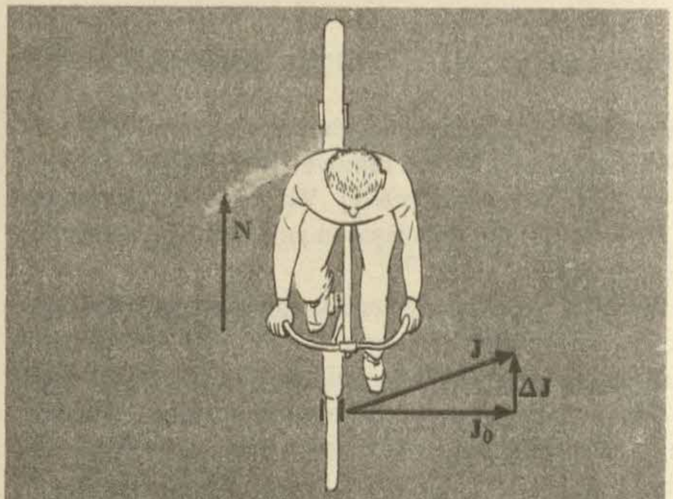
There are many examples of gyroscopic motion in daily life. Here is a student riding his bicycle, straight ahead.



A top view shows the direction of the angular momenta associated with the wheels.



To make a left turn the rider applies a torque as shown (tipping over toward *his* left.) The direction of the torque is to the rear of the bicycle . . .



and so must be the direction of ΔJ. The new J is as shown, and the bike moves off to the rider's left.

Let J denote $I_3\omega_3$; by our assumption the magnitude J is constant. The torque is in the horizontal plane and normal to the axis of the top. Let φ measure the angular position of the axis of the top in the horizontal plane. Then $\mathbf{N} = N\hat{\boldsymbol{\varphi}}$, and the equation of motion $\mathbf{J} = \mathbf{N}$ reduces to

$$\dot{\mathbf{J}} = N\hat{\boldsymbol{\varphi}}. \tag{74}$$

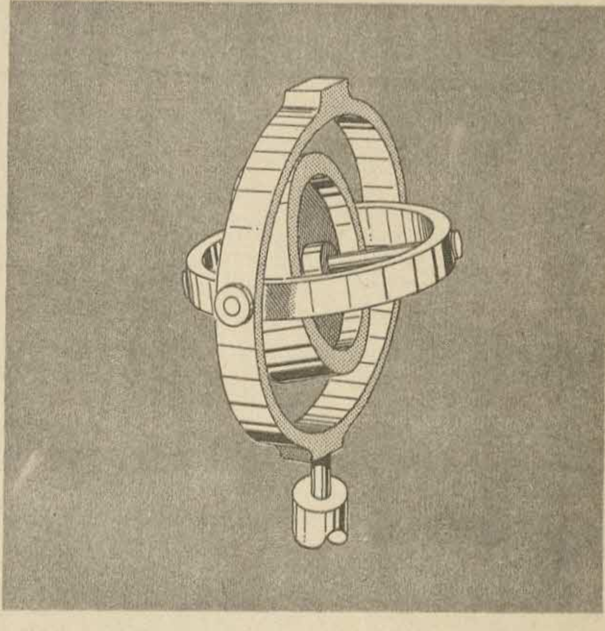
If the magnitude of $\mathbf{J}$ is to be constant, as we assumed, then a solution of (74) is

$$\dot{\mathbf{J}} = J\varphi\hat{\boldsymbol{\varphi}} = J\Omega\hat{\boldsymbol{\varphi}}. \tag{75}$$

Here Ω denotes the precessional frequency, which is the angular frequency with which the rotational axis of the gyroscope moves about the vertical axis. On combining (74) and (75) we have

$$\Omega = \frac{N}{J} = \frac{N}{I_3\omega_3}. \tag{76}$$

The directional stability of a rotating rigid body or gyroscope is best exhibited by comparing its behavior with that of a nonrotating body. Suppose both are located in force-free space. Let a slight transient torque act on both bodies. The nonrotating body will begin to rotate when the torque is applied, and the rotation will continue indefinitely after the removal of the torque. The gyroscope may deviate slightly while the torque is applied, but on removal of the torque the deviation will cease. This is due to the fact that (from $\dot{\mathbf{J}} = \mathbf{N}$) a continued torque is necessary to cause a continued change of the angular momentum vector. Expressed in another way, a small increment $\Delta\mathbf{J} = \mathbf{N}\,\Delta t$ of angular momentum looks big if compared to zero; it represents a continuous rotation about a fixed axis of a body initially nonrotating. But the same $\Delta\mathbf{J}$ added perpendicularly to the large initial $\mathbf{J}$ of the gyroscope will cause at most an angular deviation of $\Delta J/J$ in the orientation of the axis of the gyroscope. This fact is the basic explanation of the gyroscope. It has an obvious application to the operation of inertial navigation systems.



A gyroscope in a gimbal mounting (as shown in the drawing) does not experience a torque from the rotation of the earth or from any motion of the vehicle to which the mounting is attached. Thus the axis of the rotating body will always point in a fixed direction in space. We note that a symmetrical rotating body is always used in a gyroscope, in order that the rotation axis may coincide with the direction of the angular momentum.

The *gyrocompass* employs a different type of mounting. The rotation of the earth exerts a torque on the rotating body unless the rotation axis of the body points toward true north.

Film list

"Angular Momentum, A Vector Quantity" (27 min) A. Lemonick (ESI film). Shows that angular momenta add vectorially and that torque applied to a system which already has angular momentum causes the angular momentum to precess.

Problems

1. *Rotating hoop.* A circular metal hoop of mass $M = 10^3$ gm and radius $R = 20$ cm rotates about its center at 10 cycles/sec, or $2\pi(10)$ rad/sec. The axis of rotation is normal to the plane of the hoop.

(*a*) Show that the moment of inertia about this axis is 4×10^5 gm-cm^2.

(*b*) Show that the angular momentum about the same axis is 2.5×10^7 erg-sec.

(*c*) It is desired to increase the angular momentum about the same axis by 1×10^7 erg-sec. If a torque N is applied for 10 sec, show that a torque $N = 1 \times 10^6$ dyne-cm is needed.

(*d*) Suppose that this torque is applied by a force acting tangentially on the rim of the hoop; show that the value of the force is 5×10^4 dynes.

2. *Bicycle wheel.* (*a*) Make a reasonable estimate of the order of magnitude of the angular momentum of a bicycle wheel (adult size) when rolling at 30 km/hr. (First convert the velocity to cm/sec.)

Ans. $\sim 10^8$ gm-cm^2/sec.

(*b*) How much torque is needed to turn the handlebar through 1 rad in 0.1 sec? *Ans.* $\sim 10^9$ gm-cm^2/sec^2.

3. *Compton wavelength.* Suppose that an electron of mass $m \approx 1 \times 10^{-27}$ gm moves in a circular orbit of radius $R \approx 4 \times 10^{-11}$ cm. (This radius is the approximate value of the quantity $h/2\pi mc$ or $\hbar/mc$, which is a fundamental length in atomic physics known as the Compton wavelength.) At what speed (in cm/sec) must the electron travel in order to reproduce the observed value of the angular momentum, which is $\frac{1}{2}\hbar \approx \frac{1}{2} \times 10^{-27}$ erg-sec? Here $\hbar$ is (Planck's constant)/2π. You may find it convenient to take advantage of cancellations by doing the estimate in terms of symbols, starting with the angular momentum $mR^2\omega = \frac{1}{2}\hbar$ and then solving for the velocity $v = R\omega = \hbar/2mR = ?$. *Ans.* 1.5×10^{10} cm/sec.

4. *Moment of inertia of dumbbell.* A light rod of length $L = 200$ cm has a mass of 1×10^3 gm attached at each end. Consider the masses as point masses in making estimates below.

(*a*) Locate the position of the center of mass.

(*b*) What is the value of the moment of inertia about an axis perpendicular to the rod and passing through the center of mass?

(*c*) Suppose the rod lies in the xy plane of a Cartesian coordinate system, at an angle of 45° with the x axis. Evaluate the inertial coefficients I_{xx}, I_{yy}, I_{zz} about an origin at the center of the rod.

(*d*) Evaluate I_{xy}, I_{xz}, I_{yz} for the same geometry.

5. *Moment of inertia of uniform rod.* Calculate the moment of inertia of a uniform thin rod of mass M and length L:

(*a*) About an axis perpendicular to the rod at the center of mass.

(*b*) About an axis perpendicular to the rod at one end of the rod.

6. *Moment of inertia of cylinder.* Show that the moment of inertia of a uniform, solid circular cylinder (or disk) of length L, radius R, and mass M is $I = \frac{1}{2}MR^2$, taken about the longitudinal axis of the cylinder. (*Hint:* First find the moment of inertia of a thin cylindrical shell of density ρ, radius r, and thickness Δr. Integrate to obtain result for the solid cylinder.)

7. *Rolling cylinder.* What is the ratio of rotational to translational kinetic energy for a solid cylinder rolling down a plane without slipping?

Ans. 1/2.

8. *Rolling cylinder.* A solid cylinder of mass M rolls without slipping down a plane of length L inclined at an angle θ with the horizontal.

(*a*) What is the velocity of the center of mass of the cylinder at the bottom of the plane? *Ans.* $(2/\sqrt{3})(gL \sin\theta)^{1/2}$.

(*b*) If, however, the cylinder slides down the plane (without rolling), what is the final velocity?

9. *Summation convention.* In the following questions refer to Mathematical Note 2, Chap. 2, and to the discussion of the summation convention given in the text of Chap. 8.

(*a*) Use the unit vector $\hat{\mathbf{e}}_\mu$ to express the vectors **A** and **B** in terms of the components in a compact form.

(*b*) Expand the double sum $\delta_{\mu\nu}A_\mu A_\nu$, explicitly showing each term. Show that it reduces to the expansion of the expression $A_\mu A_\mu$.

(*c*) Use the results of parts (*a*) and (*b*), and the properties of the unit vectors, to show that $\mathbf{A} \cdot \mathbf{B} = A_\mu B_\mu$. The Kronecker delta and the summation convention should be utilized.

(*d*) Use the symbol $\epsilon_{\mu\nu\lambda}$ and the summation convention to express $\mathbf{A} \times \mathbf{B}$ in terms of the components of A and B. *Ans.* $\epsilon_{\mu\nu\lambda}A_\nu B_\lambda \hat{\mathbf{e}}_\mu$.

10. *Gyroscopic stabilization.* A vessel of mass 10^7 kg is gyrostabilized by a uniform circular disk of mass 5×10^4 kg and radius 2 m which rotates at 15 rev/sec.

(*a*) What is the angular momentum of the stabilizer?

The beam of the vessel is 20 m; we may consider 10 m as the effective radius of the cross section. The time of a free roll ($-20°$ to $+20°$, say) is 12 sec.

(*b*) Estimate the angular momentum of the ship in such a roll.

(*c*) By how much do you suppose the gyrostabilizer would help reduce the roll angle?

11. *Rotation of DNA molecule.* The mass of a molecule of DNA from $T2$ bacteriophage is 1.2×10^8 molecular weight. The molecule is a double helix of 1.2×10^4 turns. The radius of gyration of the helix is 6.7 Å.

(*a*) What is the moment of inertia about the axis of the helix?

Ans. $I \approx 5 \times 10^{-31}$ gm-cm^2.

(*b*) If the rotational energy $\frac{1}{2}I\omega^2$ about the axis of the helix is equal to the thermal energy $\frac{1}{2}kT$ at 300°K, what is the value of ω, the angular frequency of rotation? *Ans.* 3×10^8 rad/sec.

(*c*) If the double helix unwound into separate strands at 3×10^8 rad/sec, how long would it take to unwind? [The actual unwinding in solution proceeds in a more random way because of the random bombardment of the helix by water molecules, so that what we have calculated is very much a lower limit on the unwinding time.]

Chapter 9

Forces between a point mass and a spherical shell 263
Force between a point mass and a solid sphere 266
Gravitational and electrostatic self-energy 267
Examples: Gravitational energy of the Galaxy 268
Gravitational energy of uniform sphere 268
Fundamental lengths and numbers 270
Inverse-square-law forces and static equilibrium 274

Orbits 274
Two-body problem and the reduced mass 275
Example: Vibration of a diatomic molecule 277
One-body problem 279
Example: Circular orbit 286
Kepler's laws 288

Further reading 291

Problems 291

Advanced Topic 1: Virial theorem 296
Examples: Interior temperature of 300
the sun 301
Condensation of galaxies 302
Effect of radiative energy losses

Advanced Topic 2: Numerical calculation of an orbit 305

Inverse-Square-Law Force

The magnitudes of the electrostatic and gravitational forces between two point particles at rest are given by

$$F = \frac{C}{r^2}, \tag{1a}$$

with C a constant. Such forces are called inverse-square-law central forces. The word *central* means that the force is directed along the line which connects the two particles. If one particle is at the origin and the second particle is at position **r**, the force on the second particle is given by

$$\mathbf{F} = \frac{C}{r^2}\hat{\mathbf{r}}, \tag{1b}$$

where for the gravitational force between point masses M_1, M_2:

$$C = -GM_1M_2; \qquad G = 6.67 \times 10^{-8}\ \text{cm}^3/\text{gm-sec}^2, \tag{2}$$

and for the electrostatic force between point charges q_1, q_2:

$$C = q_1q_2, \tag{3}$$

provided that the charge is expressed in Gaussian CGS units, as in Chap. 4. The gravitational force is always attractive. The electrostatic (Coulomb) force is attractive if the charges q_1, q_2 have opposite signs, and repulsive if q_1, q_2 have the same sign.

The exponent of r in (1a) is known very accurately by experiment to be equal to $-2.000\ldots$; for electrostatic forces this is established at least down to distances of the order of 10^{-13} cm. A wide variety of experimental results would be highly sensitive to small departures from an exact inverse-square law of force. The central experiments are discussed in Vol. II with special reference to electrostatic forces. For gravitational forces we appeal for experimental support chiefly to the excellent agreement between prediction and observation of planetary motions in the solar system.

The inverse-square law of force is also expressed as an inverse-first-power law of potential energy. As we saw in Chap. 5, F is equal to $-\partial U/\partial r$. From Eq. (1*a*), then

$$F = -\frac{\partial U}{\partial r} = \frac{C}{r^2}, \tag{4a}$$

and

$$U(r) = \frac{C}{r} + \text{const.} \tag{4b}$$

If we choose $U(r)$ to be zero when the particles are infinitely far apart, then the constant of integration is zero and we have

$$U(r) = \frac{C}{r}, \tag{4c}$$

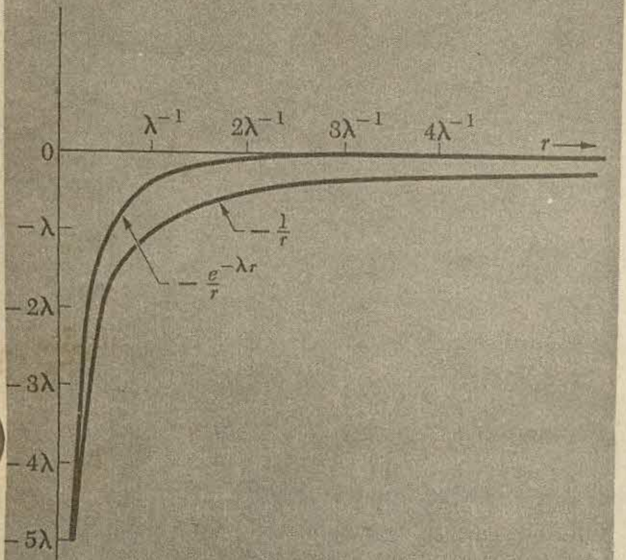
where C is given by (2) or (3) for gravitational forces or electrostatic forces respectively. Thus

$$U(r) = -\frac{GM_1M_2}{r}, \quad \text{or} \quad U(r) = \frac{q_1q_2}{r}. \tag{4d}$$

The law of attraction of two nucleons (protons or neutrons) is known from particle-scattering experiments to deviate strongly *at short distances* (very much smaller than the dimensions of an atom) from the Coulomb form e^2/r. There is an attractive nuclear interaction which gives rise to a potential energy roughly of the form

$$U_{\text{nuclear}}(r) = -D\frac{e^{-r/r_0}}{r}. \tag{5}$$

Here r_0 is a constant having the dimensions of a length; $r_0 \sim 2 \times 10^{-13}$ cm. The constant D is of the order of magnitude 10^{-18} erg-cm. Two protons (being charged) will have a Coulomb repulsion in addition, but for $r \lesssim r_0$ the nuclear interaction is the dominant interaction. We note that the exponential factor in (5) controls the range of the nuclear interaction. At a distance of 2×10^{-10} cm, which is 10^3 times r_0, the ratio of the nuclear potential energy (5) to the electrostatic potential energy e^2/r is $(D/e^2)\exp(-10^3) \approx 10^{-400}$, which is a very small number. The forces between two electrons are accurately Coulombic down to the smallest distances known, but electrons have a magnetic dipole



The attractive gravitational potential is of the form

$$-\frac{1}{r}$$

as is the attractive electrostatic potential. The function falls off slowly at great distances; the inverse-square force is thus a "long-range" force. The attractive nuclear force potential is of the form

$$-\frac{e^{-\lambda r}}{r}$$

and falls rapidly to zero at great distances.

moment in addition to their charge, and the magnetic moments of two electrons give rise to a noncentral inverse-cube law of force (Vol. II).

Given an inverse-square law of force, what special characteristics follow? In what vital respects does the universe reflect an inverse-square law? We now turn to these important questions. We shall often discuss the potential energy rather than the force. *In solving potential or force problems, the student will nearly always find it easier first to solve for the potential energy,* and then to obtain the force or force component by differentiating with respect to the separation, as in Chap. 5. The potential energy is a scalar; the force is a vector. It is simpler first to find one scalar than to find the three components of a vector.

Force Between a Point Mass and a Spherical Shell. One important consequence of the inverse-square law of force is that the force on a point test mass M_1 distant r from the center of a uniform thin spherical shell of radius R is exactly the same at points $r > R$ outside the shell as if the entire mass of the shell were concentrated at its center. A second consequence is that for points $r < R$ inside the shell the force on the point mass is *zero.* These consequences are so important that we shall give the derivations in full detail. We follow a special method of solution which takes advantage of the geometrical symmetry of the problem.

We first consider a ring on the shell having angular width $\Delta\theta$ or width $R\,\Delta\theta$, as in the accompanying figure. Let σ be the mass per unit area of the shell. The entire ring is distant r_1 from the test mass M_1. The radius of the ring is $R \sin\theta$ and the circumference is $2\pi R \sin\theta$. The area of the ring is therefore given by

$$(2\pi R \sin\theta)(R\,\Delta\theta) = 2\pi R^2 \sin\theta\,\Delta\theta. \tag{6}$$

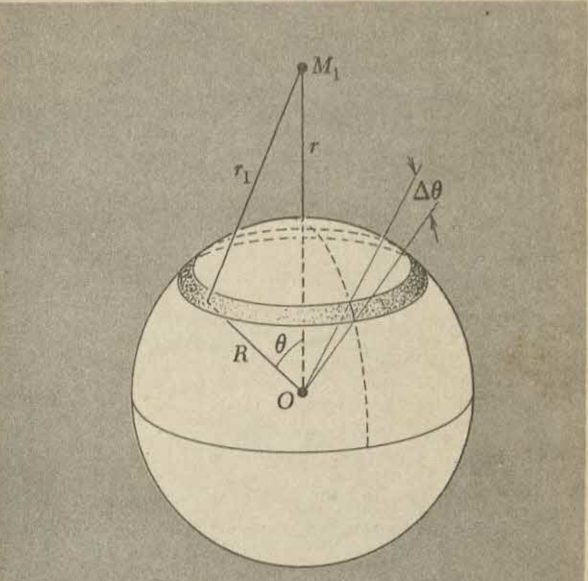
The mass of the ring is given by the product of the area times the mass per unit area σ:

$$M_{\text{ring}} = (2\pi R^2 \sin\theta\,\Delta\theta)\sigma. \tag{7}$$

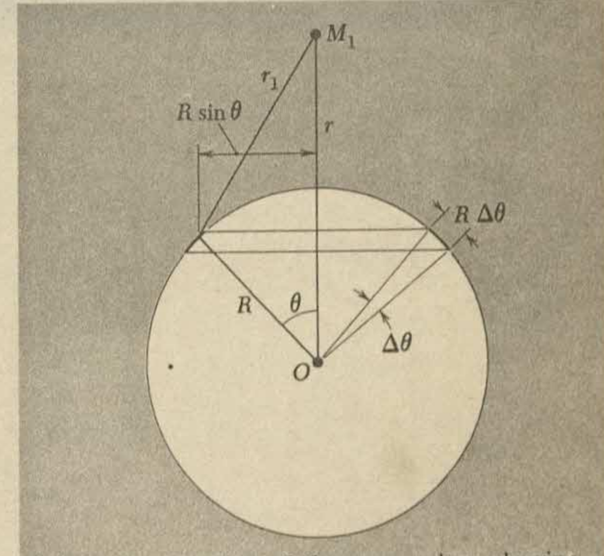
The potential energy U_{ring} of the test mass in the gravitational field of the ring is, from (7) and the relation $U = -GM_1M_{\text{ring}}/r_1$ for the gravitational potential energy between two masses,

$$U_{\text{ring}} = -\frac{GM_1(2\pi R^2 \sin\theta\,\Delta\theta)\sigma}{r_1}. \tag{8}$$

Here r_1 is the distance from the test mass to the ring.



Perspective drawing of spherical shell and point mass M_1, showing how spherical shell is divided into rings. The shell has a mass density σ per unit area.



Section drawing of the same sphere showing ring, of total area $2\pi R^2 \sin\theta\,\Delta\theta$.

All elements of the ring are equidistant from the test mass. By the law of cosines [Eq. (2.35)], applied to the triangle formed by R, r, and r_1, we have

$$r_1^2 = r^2 + R^2 - 2rR\cos\theta. \tag{9}$$

For the ring, R and r are constant because R is the radius of the shell and r the distance from the center of the shell to the test mass M_1. Under these conditions the differential of (9) is

$$2r_1\,dr_1 = -2rR\,d(\cos\theta) = 2rR\sin\theta\,d\theta. \tag{10}$$

This useful relation enables us to rewrite (8) to obtain

$$U_{\text{ring}} = -\frac{GM_1(2\pi R\,\Delta r_1)\sigma}{r}. \tag{11}$$

Notice that in the denominator there occurs the distance r from the test mass to the center of the sphere.

Now the total potential energy U_{shell} of the test mass in the gravitational field of the spherical shell is given by the sum of U_{ring} over all the rings which make up the shell. When we sum, we have only to sum over Δr_1. When the test mass lies outside the shell, the range of values of r_1 is seen to run from $r - R$ to $r + R$, so that

$$\Sigma\,\Delta r_1 = (r + R) - (r - R) = 2R. \tag{12}$$

It is very fortunate that the problem can be reduced to such a simple summation. Using (12) to sum over (11) we have

$$U_{\text{shell}} = \Sigma\,U_{\text{ring}} = -\frac{GM_1 2\pi R\sigma}{r}\Sigma\,\Delta r_1 = -\frac{GM_1 4\pi R^2\sigma}{r}. \tag{13}$$

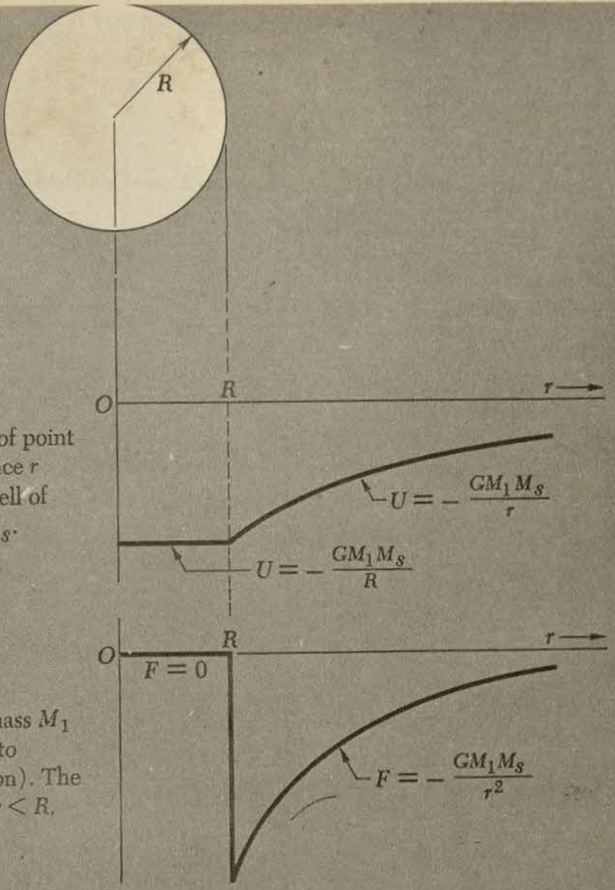
But $4\pi R^2$ is the surface area of the spherical shell, and $4\pi R^2\sigma$ is the mass M_s of the shell. We can therefore rewrite (13) as

$$U_{\text{shell}} = -\frac{GM_1M_s}{r} \qquad (r > R), \tag{14}$$

where r is the distance between the test mass and the center of the spherical shell. We have shown that *the spherical shell acts at points outside as if all its mass M_s were concentrated at the center of the shell.*

If the test mass lies anywhere *within* the shell, the range of summation of Δr_1 in $\Sigma\,U_{\text{ring}}$ is from $R - r$ to $R + r$, so that now

$$\Sigma\,\Delta r_1 = (R + r) - (R - r) = 2r. \tag{15}$$



Shell of radius R, mass M_s.

Potential energy of point mass M_1 at distance r from center of shell of radius R, mass M_s.

Force on point mass M_1 (negative sign to indicate attraction). The force is zero for $r < R$.

Using (15) to sum over (11) we obtain

$$U_{\text{shell}} = \Sigma\, U_{\text{ring}} = -\frac{GM_1 2\pi R\sigma}{r}\,\Sigma\,\Delta r_1$$

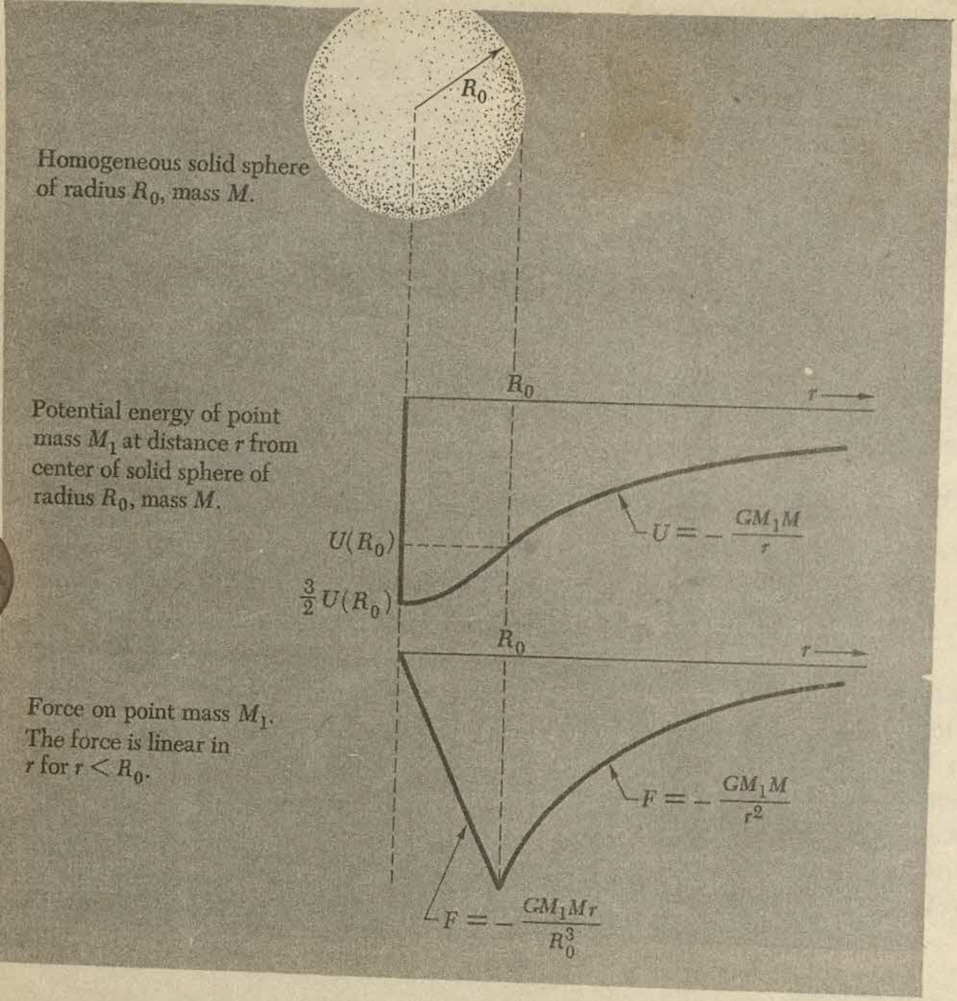
$$= -\,GM_1 4\pi R\sigma = -\frac{GM_1 4\pi R^2\sigma}{R}$$

$$= -\frac{GM_1M_s}{R} \qquad (r < R). \tag{16}$$

The potential (16) is constant at all points in the interior of the shell, and is equal to (14) evaluated at $r = R$.

We have seen before that the magnitude of the force F on the test mass M_1 is equal to $-\partial U/\partial r$, because the force is in the radial direction. From (14) and (16) we have for the force due to the shell:

$$F = -\frac{\partial U}{\partial r} = \begin{cases} -\dfrac{GM_1M_s}{r^2} & (r > R); \\ 0 & (r < R). \end{cases} \tag{17}$$


R_0
Homogeneous solid sphere of radius R_0, mass M.
Potential energy of point mass M_1 at distance r from center of solid sphere of radius R_0, mass M.
R_0
$r \rightarrow$
$U(R_0)$
$\frac{3}{2}U(R_0)$
$U = -\frac{GM_1M}{r}$
Force on point mass M_1. The force is linear in r for $r < R_0$.
R_0
$r \rightarrow$
$F = -\frac{GM_1M}{r^2}$
$F = -\frac{GM_1Mr}{R_0^3}$

Thus there is no force on a test mass inside the shell. This is a very special property of an inverse-square-law force. Outside the shell the force varies as $1/r^2$, where r is measured from the center of the shell.

Force Between a Point Mass and a Solid Sphere. We may build up a solid sphere of mass M and radius R_0 by adding up a series of concentric shells. For points r outside the sphere we have, using (14), the following result for the potential energy of the test mass M_1 in the gravitational field of the solid sphere:

$$U_{\text{sphere}} = \Sigma\, U_{\text{shell}} = -\frac{GM_1}{r}\Sigma M_s = -\frac{GM_1M}{r}. \qquad (18)$$

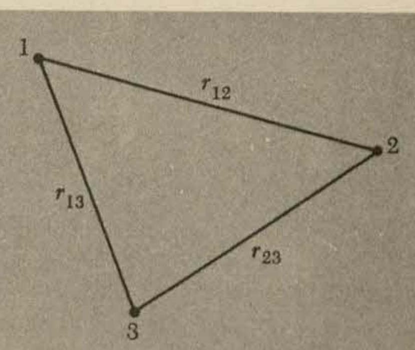
Recall that r is the distance from the test mass to the center of the sphere.

The magnitude of the force on M_1 is, for $r > R_0$,

$$\boxed{F = -\frac{\partial U}{\partial r} = -\frac{GM_1M}{r^2}.} \qquad (19)$$

This is the central result of our analysis. It could have been obtained also by a direct integration of the force components over the shell, but the mathematics is more concise as we have done the problem. By an easy extension of (19), we see that the force between two uniform spheres of masses M_1, M_2 is equal to the force between two point masses M_1, M_2 at their respective centers. Having replaced one sphere by a point mass, we can then replace the second sphere by a point mass. This happy result simplifies many calculations.

Gravitational and Electrostatic Self-energy. The self-energy of a body is defined to be the work done in assembling the body from infinitesimal elements that are initially an infinite distance apart. Let us consider the gravitational self-energy; this will have a negative sign, because the gravitational force is attractive. (We have to do positive work against gravity to separate the atoms of a star, taking each atom to infinity.) We need the gravitational self-energy usually for stellar and galactic problems. The electrostatic self-energy is often calculated for crystals, both insulators and metals.



The gravitational potential energy of three atoms of masses M_1, M_2, M_3 is

$$U = -G\left(\frac{M_1M_2}{r_{12}} + \frac{M_1M_3}{r_{13}} + \frac{M_2M_3}{r_{23}}\right).$$

The potential energy of N discrete atoms due to their mutual gravitational attraction is equal to the sum of the potential energy of all pairs of atoms:

$$U_s = -G \sum_{\substack{\text{All}\\ \text{pairs}\\ i\neq j}} \frac{M_iM_j}{r_{ij}} = -G \sum_{i>j}^{N} \sum_{j=1}^{N} \frac{M_iM_j}{r_{ij}}. \qquad (20)$$

We have written the result in two forms. In the form in the middle of the line we indicate that the summation is to be carried out over all pairs of atoms i and j, excluding the case $i = j$, which is not a pair at all. The case $i = j$ could only relate to the gravitational self-energy of a single atom, and we usually want to leave this out because we believe it is not changed on assembling the atom into a star. The form on the right-hand side is just another way of indicating that we want to count each pair ij once: thus we count 4,3 and not 3,4. We could equally well write

$$U = -\frac{1}{2} G \sum_{i=1}^{N}{}' \sum_{j=1}^{N} \frac{M_iM_j}{r_{ij}}. \qquad (21)$$

Here we count each pair twice, once for example as 3,4 and once as 4,3. The factor $\frac{1}{2}$ corrects for the overcounting. (This factor of

$\frac{1}{2}$ tends to appear in self-energy problems.) The prime on the left-hand summation Σ is a coded instruction to the reader to omit the terms $i = j$ from the summation.

Example. Gravitational energy of the Galaxy. Let us estimate the gravitational energy of the Galaxy. If we omit from the calculation the gravitational self-energy of the individual stars, then we need only estimate the value of the expression (21).

We approximate the gross composition of the Galaxy by N stars, each of mass M, and with each pair of stars at a mutual separation of the order of R. Then (21) reduces to

$$U \approx -\frac{1}{2}G(N-1)N\frac{M^2}{R}. \qquad (22)$$

(There are N equal terms in $\sum_{j=1}^{N}$ and $(N-1)$ terms in $\sum_{i=1}^{N}{}'$). If $N \approx 1.6 \times 10^{11}$; $R \approx 10^{23}$ cm; and $M \approx 2 \times 10^{33}$ gm (as for the sun), then

$$\begin{aligned} U &\approx -\tfrac{1}{2}(7 \times 10^{-8})(1.6 \times 10^{11})^2(2 \times 10^{33})^2/10^{23} \\ &\approx -4 \times 10^{58} \text{ ergs}. \end{aligned} \qquad (23)$$

Example. Gravitational energy of uniform sphere. It is not difficult to evaluate the self-energy U_s of a uniform sphere of mass M and radius R. We convert the double summations in (21) to integrals and then carry out the multiple integrations. But let us first attempt to guess the answer. What would we expect it to be? The answer must involve G, M, and R, in the correct dimensions. Why not

$$U_s \approx -\frac{GM^2}{R} \quad ? \qquad (24)$$

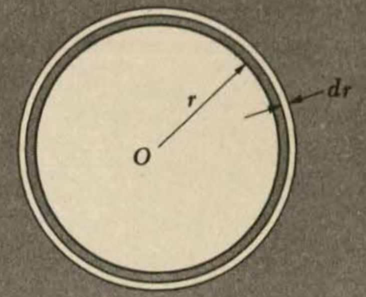
This is in fact correct, except for a numerical factor of the order of unity.

When the calculation is done exactly, we find for the uniform sphere the result

$$\boxed{U_s = -\frac{3}{5}\cdot\frac{GM^2}{R},} \qquad (25)$$

so that our estimate (24) is pretty good. To obtain the factor $\frac{3}{5}$, we build up the solid sphere in a special way. We first consider the energy of interactions between a solid spherical core of radius r and a surrounding spherical shell of thickness dr. If ρ denotes the density, the mass of the core is $(4\pi/3)r^3\rho$ and the mass of the shell is $(4\pi r^2)(dr)\rho$. Thus the gravitational potential energy of the shell due to the presence of the core is, from (14),

$$\frac{-G\left(\frac{4\pi}{3}r^3\rho\right)\left(4\pi r^2\,dr\,\rho\right)}{r} = -\frac{1}{3}G(4\pi\rho)^2 r^4\,dr. \qquad (26)$$



Spherical shell of thickness dr surrounding a solid spherical core of radius r. By adding successive shells we construct a solid sphere of radius R. The area of one surface of the shell is $4\pi r^2$; the thickness is dr, so that the volume of the shell is $4\pi r^2\,dr$.

The self-energy of the solid sphere is given by the integral of (26) between $r = 0$ and $r = R$. The integration corresponds to adding successive shells to the core until the core has radius R. Initially the core has zero radius. On integrating (26), we obtain the result

$$U_s = -\frac{1}{3}G(4\pi\rho)^2\cdot\frac{1}{5}R^5 = -\frac{3}{5}G\left(\frac{4\pi}{3}\rho R^3\right)^2\cdot\frac{1}{R}$$
$$= -\frac{3}{5}\cdot\frac{GM^2}{R}, \qquad (27)$$

because the mass is given by

$$M = \frac{4\pi}{3}\rho R^3. \qquad (28)$$

The gravitational self-energy of the sun is, from (27), with $M_s \approx 2\times10^{33}$ gm and $R_s \approx 7\times10^{10}$ cm,

$$U_s \cong -\frac{(3)(7\times10^{-8})(2\times10^{33})^2}{(5)(7\times10^{10})} \cong -2\times10^{48}\text{ ergs}. \qquad (29)$$

This is a lot of energy! The sun may complete its evolution as a dense white dwarf star, with a radius about 0.1 of its present radius. It is apparent that a large amount of gravitational energy will be released in the contraction.

The electrostatic self-energy of a uniform spherical distribution of total charge e and radius R is obtained by substituting e^2 for $-GM^2$ in the result (25):

$$U_s = \frac{3}{5}\cdot\frac{e^2}{R}. \qquad (30)$$

To estimate the electrostatic self-energy of an electron, we need to know the radius R. Because we do not have a fundamental theory of the electron, all we can do is to proceed backward and estimate the radius from the energy.

There is a famous relation due to Einstein which says that a mass M is always associated with an energy E according to the equation

$$E = Mc^2, \tag{31}$$

where c is the velocity of light. (We derive this in Chap. 12.) If the energy of the electron were entirely the electrostatic energy of a uniform charge distribution, then we would have

$$U_s = \frac{3e^2}{5R} = mc^2, \tag{32}$$

which would determine the radius of the electron. But we do not know the structure of the electron in detail. The model we have outlined cannot be entirely satisfactory, for what keeps the charge in the electron together? Why doesn't it fly apart under the Coulomb repulsion of like-charge elements? At present we have no theory of why there is an electron.

So let us drop the factor $\frac{3}{5}$ in (32). It would be pretentious to keep the factor because it suggests a refinement of knowledge about the electron that we do not possess. We define (by universal convention) a length r_0 by the relation

$$\frac{e^2}{r_0} \equiv mc^2; \qquad r_0 \equiv \frac{e^2}{mc^2} = 2.82 \times 10^{-13}\ \text{cm}. \tag{33}$$

This length is called the *classical radius of the electron*. It has something to do with the electron, but we don't know exactly what! Nevertheless, it is called a fundamental length. We now will have a short digression on the love of physicists for numbers.

Fundamental Lengths and Numbers. Numerology is not physics, any more than astrology is astronomy. But numbers play an important role in physics. When we see that constants which have something to do with a problem (as e, m, and c have something to do with electromagnetism and the electron) may be combined to form a length, we may wonder just what that length means, if anything. It is a legitimate and very provocative question to ask. Some fundamental quantities have an obvious meaning, others do not.

Dimensionless numbers are especially interesting. They stimulate thinking. In hydrodynamics there occurs a dimensionless number called Reynolds' number. When Reynolds' number is large, it is observed that the flow of liquid is often turbulent; when small, the flow is nonturbulent or laminar. In atomic physics we can form an important dimensionless number from e, h, and c. Here h is Planck's constant; we prefer to deal with $\hbar = h/2\pi$. Planck's constant is defined by the relation $E = h\nu$ for light waves; it gives the connection between the frequency ν and the energy E of a photon. Thus h (and $\hbar$) have the dimensions of energy × time. We know that e^2/r_0 is an energy, so that e^2 has the dimensions of energy × length.

If we divide e^2 by $\hbar$ we get a quantity having the dimensions of length/time, or velocity. An obviously important velocity is c, the velocity of light. If we divide $e^2/\hbar$ by c, we have the dimensionless quantity denoted by α:

$$\alpha \equiv \frac{e^2}{\hbar c} = \frac{1}{137.04}. \tag{34}$$

This quantity is called the *fine-structure constant.* The reasons for the name are historical, having to do with the splitting of spectral lines. We do not know why $e^2/\hbar c$ has this particular value, nor even whether it is something which can ever be deduced from any theory. It is discussed in Vol. IV.

We can express a whole series of important fundamental lengths by dividing the classical radius of the electron by powers of α. One important length often seen in quantum physics is the *Compton wavelength* $\lambdabar_C$ of an electron:[†]

$$\lambdabar_C = \frac{r_0}{\alpha} = \frac{\hbar c}{e^2}\cdot\frac{e^2}{mc^2} = \frac{\hbar}{mc} = 3.86 \times 10^{-11}\text{ cm}. \tag{35}$$

Sometimes the Compton wavelength of the electron is defined as $\lambda_C \equiv 2\pi\lambdabar_C$.

[†] The symbol λ is the Greek character lambda; the symbol $\lambdabar$ is read as "lambda-bar" and is defined as $\lambda/2\pi$.

Another important length is the *Bohr radius of the ground state of hydrogen:*

$$a_0 = \frac{r_0}{\alpha^2} = \frac{\lambda\!\!\!^{-}_c}{\alpha} = \frac{\hbar c}{e^2} \cdot \frac{\hbar}{mc} = \frac{\hbar^2}{e^2 m}$$
$$= 0.529 \times 10^{-8} \text{ cm.} \tag{36}$$

This is essentially the radius of the hydrogen atom. If a_0 is related to the separation between the proton and electron in the hydrogen atom, and if the electron is bound to the proton by an electrostatic interaction, then we might guess that the binding energy (ionization energy) should be of the order of magnitude of e^2/a_0, which is about 27 ev. The actual binding energy is found by experiment and by detailed calculation to be $e^2/2a_0$.

The longer your experience in physics or engineering or astronomy or chemistry, the more you will appreciate the importance of the physical constants. We consider another example. It is natural to consider the length defined when the gravitational self-energy of a body is set equal to Mc^2:

$$\frac{GM^2}{R_0} = Mc^2;$$
$$R_0 = \frac{GM}{c^2}. \tag{37}$$

We may call R_0 the *gravitational length.* Is it relevant to anything?

In the table of values inside the covers you will find 10^{80} given as an estimate of the number of nucleons (protons or neutrons) in the universe. The mass of a nucleon is about 10^{-24} gm, so that the mass of the universe is about 10^{56} gm. Thus, from (37),

$$R_0 \approx \frac{(10^{-7})(10^{56})}{(3 \times 10^{10})^2} \approx 10^{28} \text{ cm,} \tag{38}$$

which agrees with the listing given in the table of values for the radius of the universe. This agreement might mean only that whoever made the original estimate of the mass of the universe used this dimensional relation to estimate the mass (the radius can be estimated independently). There are, however, ways of estimating by observation lower and upper limits on the mass of the universe. We believe 10^{56} gm to be close to the correct order of magnitude. (The general theory of relativity does suggest that there is a fundamental significance to the relation by which we have defined R_0. It turns out that a light signal (photon) cannot escape from the surface of a mass whose radius R is *less* than R_0. Thus an object with $R \leqq R_0$ cannot be luminous and would be invisible.)

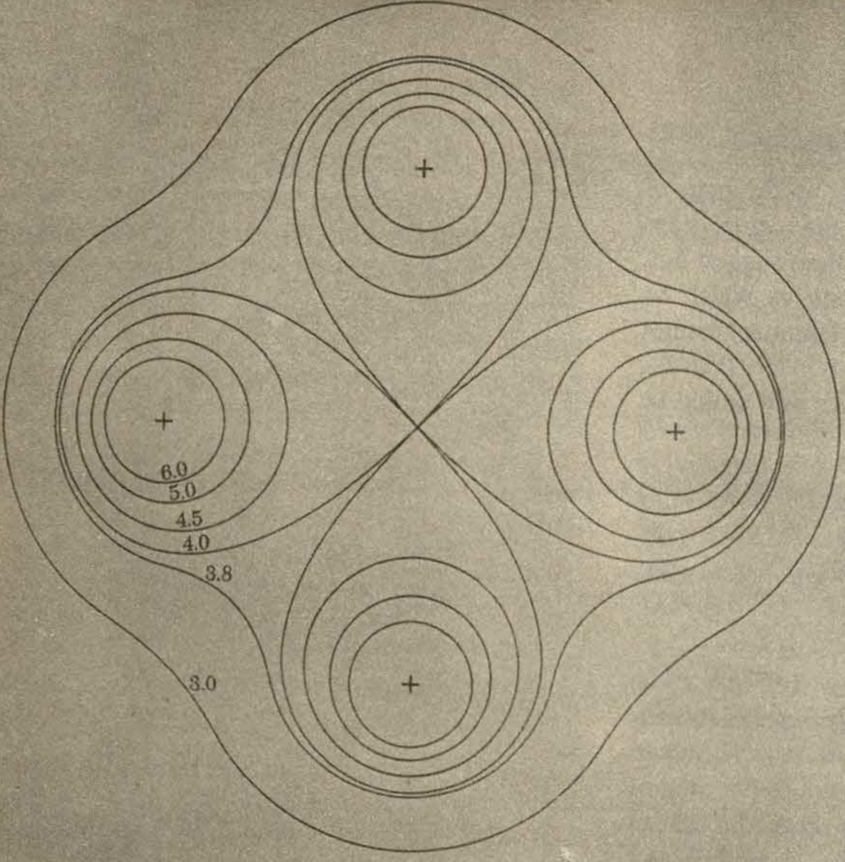
If we take M as the mass of the sun, then

$$R_0 \approx \frac{(7 \times 10^{-8})(2 \times 10^{33})}{(3 \times 10^{10})^2} \approx 10^5 \text{ cm}, \tag{39}$$

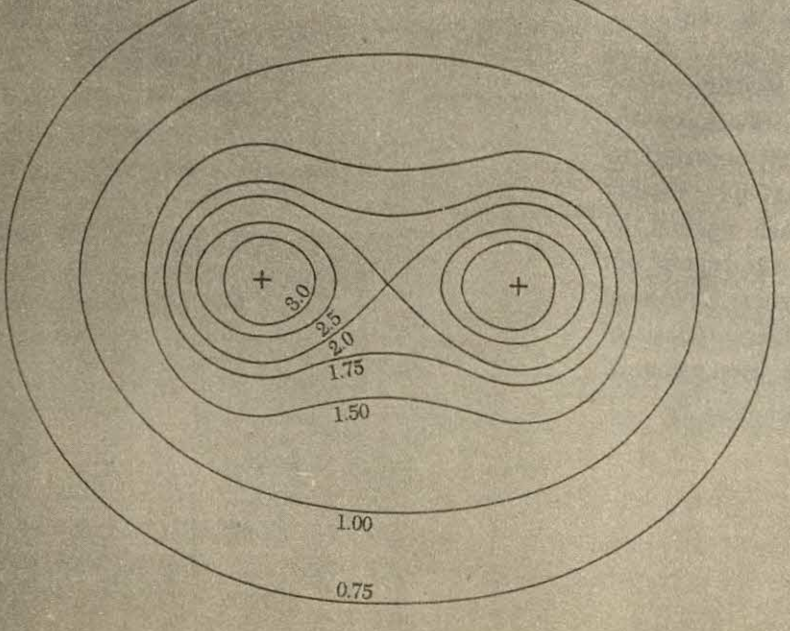
which is very much less than the radius $R = 7 \times 10^{10}$ cm of the sun. Now R_0 can only affect a problem when it is large, not when it is small (it would be zero if $G = 0$). Because $R \gg R_0$, it can have nothing much to do with the sun, at least not in the present state of the sun. Later, when the sun has nearly burned itself out and has contracted, R_0 may become important. A white dwarf star may have a mass equal to that of the sun and a radius equal to that of the earth (6×10^8 cm). But we see from (39) that even in a white dwarf the gravitational length is very small in comparison with the radius. Recently, however, there has been some speculation among astronomers that stars (neutron stars) with dimensions comparable to their gravitational length do exist.

We can form other fundamental quantities with the dimensions of time (or frequency), mass, velocity, etc. When solving a problem in a particular field of science, it is an excellent approach to form and to evaluate the fundamental quantities that may possibly be relevant. These can be helpful warning signals against the neglect of effects which may be unimportant in one regime but dominant in another. Bridge and airplane designers have seen catastrophic effects arise from an occasional underemphasis of effects whose approximate magnitude could have been obtained by calculations on the back of an envelope.

Inverse-square-law Forces and Static Equilibrium. In Vol. II we show that there can be no stable *static* equilibrium among a group of point masses (or point charges) interacting with inverse-square-law forces. By static we mean with all masses at rest. This result is made plausible by the accompanying figures, which show lines of equal values of the potential function due to two and four equal positive masses at fixed positions. A test mass placed at the center of the diagram will run away toward one or the other of the positive charges.



Contour map of equipotential surfaces between four equal masses.



Contour map of equipotential surfaces between two equal masses.

Orbits

We now consider the Kepler problem: to find the orbits of two particles interacting with inverse-square forces. The classical example is the solar system; the motion of satellites about a planet and the motion of double stars about each other are also important examples. The equation of motion $\mathbf{F} = M\mathbf{a}$ for the ith particle of a system of N particles is

$$M_i \frac{d^2\mathbf{r}_i}{dt^2} = -GM_i \sum_{j=1}^{N}{}' \frac{M_j}{{r_{ij}}^2}\hat{\mathbf{r}}_{ij}. \quad (40)$$

The prime on the Σ means that we are to omit the term $j = i$ from the summation. In (40)

$$\mathbf{r}_{ij} \equiv \mathbf{r}_i - \mathbf{r}_j \quad (41)$$

is the vector from particle j to particle i. The terms on the right-hand side of (40) are the gravitational forces exerted on particle i by the $N - 1$ other particles. The equation is written in vector form and thus it represents three separate equations for the three components of $\mathbf{r}_i$. For N particles we must solve $3N$ equations simultaneously, and these are subject to $6N$ initial conditions on the $3N$ coordinates and $3N$ velocities taken together. This does not sound very easy.

If the number of particles is not large, it is easy to solve the equations by numerical methods, either with a desk calculator or with an electronic computer. Numerical methods are standard for solving orbit problems whenever more than two particles are present. When the particles are uniform spheres, the solution of the two-particle problem can be given in analytic form; we shall develop the general analytic solution of the two-particle problem. Exact analytic solutions, which are rare in physics, are elegant, but may have no more intrinsic value than do numerical solutions. One should not underestimate the ease and power of numerical computational methods. In Advanced Topic 2, at the end of this chapter, we show an example of a numerical calculation.

Two-body Problem and the Reduced Mass. Two-body problems with central forces can always be reduced to the form of one-body problems. This is a great simplification. Although the solution for the orbit still involves a substantial number of steps, the reasoning is not difficult. The equations of motion (in a single inertial frame) of two uniform spherical bodies attracted to each other by a gravitational force are

$$M_1 \frac{d^2\mathbf{r}_1}{dt^2} = -\frac{GM_1M_2}{r^2}\hat{\mathbf{r}}; \qquad M_2 \frac{d^2\mathbf{r}_2}{dt^2} = \frac{GM_1M_2}{r^2}\hat{\mathbf{r}}. \quad (42)$$

Here $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$. The sum of the two equations (42) is

$$M_1\ddot{\mathbf{r}}_1 + M_2\ddot{\mathbf{r}}_2 = 0. \quad (43)$$

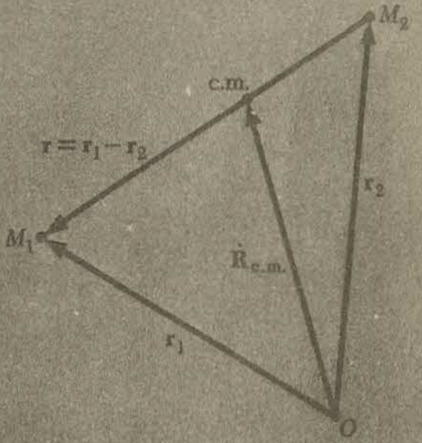
This is just a statement of the conservation of the linear momentum of the combined system, for on integrating (43) with respect to time we have

$$M_1\dot{\mathbf{r}}_1 + M_2\dot{\mathbf{r}}_2 = \text{const.} \quad (44)$$

The left-hand side is the total linear momentum.

The position of the center of mass of the two bodies is given by

$$\mathbf{R}_{\text{c.m.}} = \frac{M_1\mathbf{r}_1 + M_2\mathbf{r}_2}{M_1 + M_2}. \quad (45)$$


M_2
c.m.
$\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$
$\mathbf{r}_2$
M_1
$\dot{\mathbf{R}}_{\text{c.m.}}$
$\mathbf{r}_1$
O

M_1 and M_2 interact via central force collinear with vector $\mathbf{r}$; $\mathbf{r}_1$ and $\mathbf{r}_2$ are the position vectors of M_1 and M_2 referred to some inertial frame with origin O. In the absence of external forces, $\dot{\mathbf{R}}_{\text{c.m.}} = \text{const.}$

On differentiation of both sides of (45) with respect to time we have

$$\dot{\mathbf{R}}_{\text{c.m.}} = \frac{M_1\dot{\mathbf{r}}_1 + M_2\dot{\mathbf{r}}_2}{M_1 + M_2}; \tag{46}$$

by (44) we see that $\dot{\mathbf{R}}_{\text{c.m.}}$ = const. Thus the center of mass moves with a constant velocity. We may always take this velocity to be zero by choosing an appropriate inertial frame of reference.

Now rewrite the equations of motion (42) as

$$\frac{d^2\mathbf{r}_1}{dt^2} = -\frac{1}{M_1}\frac{GM_1M_2}{r^2}\hat{\mathbf{r}}; \qquad \frac{d^2\mathbf{r}_2}{dt^2} = \frac{1}{M_2}\frac{GM_1M_2}{r^2}\hat{\mathbf{r}}. \tag{47}$$

We subtract one equation from the other in (47) to obtain,

$$\frac{d^2(\mathbf{r}_1 - \mathbf{r}_2)}{dt^2} = \frac{d^2\mathbf{r}}{dt^2} = -\left(\frac{1}{M_1} + \frac{1}{M_2}\right)\frac{GM_1M_2}{r^2}\hat{\mathbf{r}}. \tag{48}$$

This equation involves only the single vector $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$. We can simplify the appearance of (48) by introducing the *reduced mass* μ:

$$\boxed{\frac{1}{\mu} \equiv \frac{1}{M_1} + \frac{1}{M_2}.} \tag{49}$$

Now (48) assumes the form

$$\boxed{\mu\frac{d^2\mathbf{r}}{dt^2} = -\frac{GM_1M_2}{r^2}\hat{\mathbf{r}}.} \tag{50}$$

This is a *one-body* problem: we need only solve for the motion of the single vector $\mathbf{r}$ as a function of time. The original two-body problem in Eq. (42) involved both $\mathbf{r}_1$ and $\mathbf{r}_2$.

We use (49) and (50) in the following way: Recall that $\mathbf{r}$ is the vector from M_2 to M_1. With Eq. (50) we may solve for the motion of M_1 relative tc M_2, exactly *as though* M_2 were the fixed origin of an inertial frame, except that we must use μ instead of M_1 as the mass. We have thus reduced the two-body problem to a one-body problem involving the motion of a body of mass μ. But note that the force in (50) is *not* $-G\mu M_2/r^2$! To find the orbit for the two-body problem we need only solve this one-body problem. The reduction of the two-body problem to a one-body problem can be accomplished in the same way for any central force; the reduced mass will always appear.

The reduced mass must have a value less than either M_1 or M_2. For $M_1 = M_2 = M$,

$$\frac{1}{\mu} = \frac{2}{M}; \qquad \mu = \frac{1}{2}M. \tag{51}$$

If $M_1 \ll M_2$, we have from (49),

$$\mu = \frac{M_1 M_2}{M_1 + M_2} = M_1 \frac{1}{(M_1/M_2) + 1} \cong M_1\left(1 - \frac{M_1}{M_2}\right). \tag{52}$$

We have expanded the fraction by the binomial theorem and retained only the term of lowest order in M_1/M_2. If M_1 is m (the mass of the electron), and M_2 is M_p (the mass of the proton), the reduced mass is

$$\mu \cong m\left(1 - \tfrac{1}{1836}\right). \tag{53}$$

The lighter of the two masses tends to dominate the value of the reduced mass. The departure of μ from m is easily detectable in the spectrum of atomic hydrogen.

Positronium is a hydrogenlike atom made up of a positron and an electron, with no proton. A positron is a particle which has a mass equal to the electron mass but has positive charge e. The result (51) suggests correctly that there may be a similarity between the line spectra of atomic hydrogen and of positronium, provided we make allowance for the fact that the reduced mass of positronium is about one-half that of atomic hydrogen. The Coulomb interaction between an electron and a positron has the same form as between an electron and a proton.

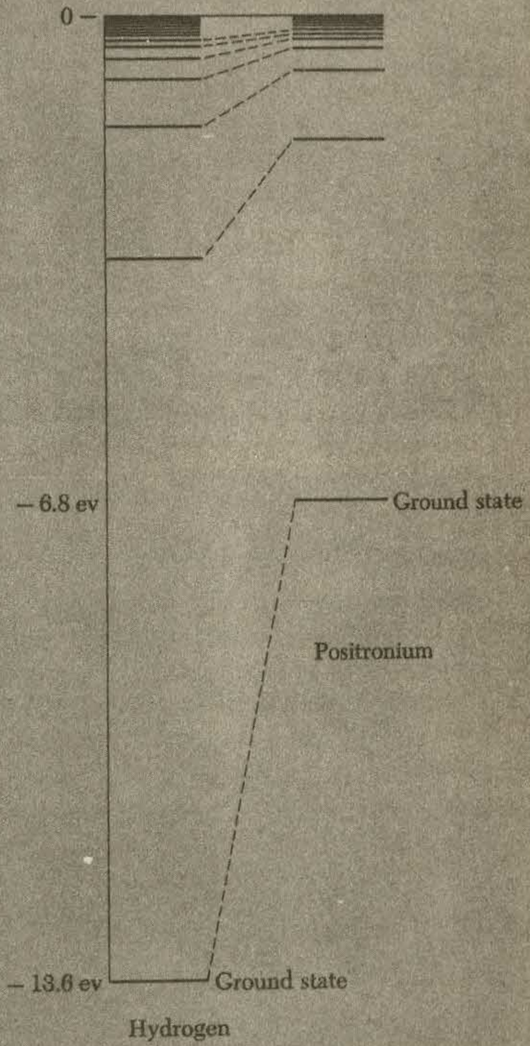
Example: Vibration of a diatomic molecule. Two atoms bound together as a stable molecule will have a potential energy which is a quadratic function of the difference $\mathbf{r} - \mathbf{r}_0$ of their separation $\mathbf{r}$ from the equilibrium separation $\mathbf{r}_0$:

$$U(r) = \tfrac{1}{2}C(r - r_0)^2, \tag{53a}$$

provided $(r - r_0)/r_0 \ll 1$. The force is along the line connecting the atoms and is given by (if the molecule is *not* rotating)

$$F = -\frac{dU}{d(r - r_0)} = -C(r - r_0). \tag{53b}$$

This describes a harmonic oscillator of force constant C. The masses of the atoms are M_1 and M_2. What is the frequency of vibration?



Energy levels of the hydrogen atom and of the positronium atom. The reduced mass for hydrogen is

$$\mu = \frac{m_e}{1 + \frac{1}{1836}} \cong m_e.$$

The reduced mass for positronium is

$$\mu = \tfrac{1}{2}m_e;$$

this causes a difference of a factor of 2 in the energies.

In a free vibration both atoms will be in motion while the center-of-mass remains at rest. The equation of motion is naturally given by (50) with (53*b*) substituted for the gravitational force:

$$\mu \frac{d^2\mathbf{r}}{dt^2} = -C(r - r_0)\hat{\mathbf{r}}. \tag{53c}$$

If the molecule is not rotating, the direction of **r** is fixed and thus

$$\frac{d^2\mathbf{r}}{dt^2} = \frac{d^2r}{dt^2}\hat{\mathbf{r}}. \tag{53d}$$

(The derivative of **r** is not this simple if the direction of **r** is changing—compare (58) below.) With (53*d*) we may rewrite (53*c*) as

$$\mu \frac{d^2r}{dt^2} = -C(r - r_0), \tag{53e}$$

which is the equation of motion of a simple harmonic oscillator of angular frequency

$$\omega_0 = \left(\frac{C}{\mu}\right)^{1/2}. \tag{53f}$$

It is known from spectroscopic measurements that the fundamental vibrational frequencies of the molecules HF and HCl are

$$\omega_0(\mathrm{HF}) = 7.55 \times 10^{14} \text{ rad/sec};$$
$$\omega_0(\mathrm{HCl}) = 5.47 \times 10^{14} \text{ rad/sec}.$$

Let us use these data to compare the force constants C_{HF} and C_{HCl}. The reduced mass of HF is, in atomic mass units,

$$\frac{1}{\mu_{\mathrm{HF}}} \cong \frac{1}{1} + \frac{1}{19} = \frac{20}{19}; \qquad \mu_{\mathrm{HF}} \cong 0.950.$$
$$\frac{1}{\mu_{\mathrm{HCl}}} \cong \frac{1}{1} + \frac{1}{35} = \frac{36}{35}; \qquad \mu_{\mathrm{HCl}} \cong 0.973.$$

(Here we have used the atomic mass of the most abundant isotope of chlorine, Cl^{35}.) Notice that the reduced masses are quite close to each other in value. This is because the hydrogen, being lightest, does most of the oscillating.

Now from (53*f*) we have for the ratio of the force constants:

$$\frac{C_{\mathrm{HF}}}{C_{\mathrm{HCl}}} = \frac{(\mu\omega_0{}^2)_{\mathrm{HF}}}{(\mu\omega_0{}^2)_{\mathrm{HCl}}} \cong \frac{54.0 \times 10^{28}}{29.0 \times 10^{28}} \cong 1.86, \tag{53g}$$

while for an individual force constant

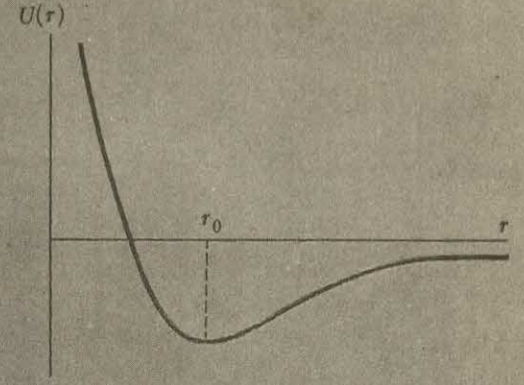
$$C_{\mathrm{HF}} \cong (54 \times 10^{28})(1.66 \times 10^{-24}) \cong 9.0 \times 10^{5}\ \mathrm{dyne/cm}; \tag{53h}$$

here we have inserted the factor which converts the mass from atomic mass units to grams.

Is this value of C reasonable? Suppose we stretch the molecule (which is about 1 Å or 1×10^{-8} cm in length) by 0.5 Å. The work needed to do this would probably be nearly enough to break up the molecule into separate atoms of H and F. By (53*a*) the work to stretch 0.5 Å should be of the order of magnitude

$$\tfrac{1}{2}C(r - r_0)^2 \approx \tfrac{1}{2}(9 \times 10^{5})(0.5 \times 10^{-8})^2 \approx 1 \times 10^{-11}\ \mathrm{erg},$$

or $\approx (1 \times 10^{-11})/(1.6 \times 10^{-12}) \approx 6$ ev. This is not unreasonable for an energy of decomposition into separate atoms. In making this estimate we have used the form (53*e*) beyond the region in which it is valid. The actual intermolecular potential energy will be more of the form shown in the accompanying figure.



Graph of the potential energy as a function of distance between two atoms combined in a molecule. The position of equilibrium is at r_0.

One-body Problem. The two-body problem for uniform spheres or point particles has been reduced to the solution of the one-body problem defined by (50):

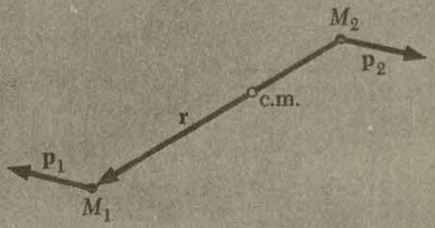
$$\mu \frac{d^2\mathbf{r}}{dt^2} = -\frac{GM_1M_2}{r^2}\,\hat{\mathbf{r}}. \tag{54}$$

Here the vector $\mathbf{r}$ connects the two original particles. The motion of the center of mass of the two particles does not involve the interparticle force but is determined by any external forces which may be present. Let us concentrate on the relative motion of the two particles, as determined by (54). This is the most important problem in classical mechanics. The solution utilizes about everything we know plus a few tricks.

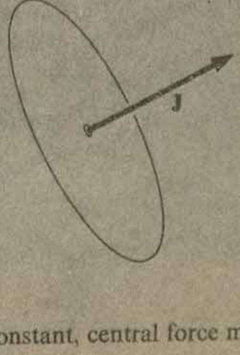
We know from Chap. 6 that the angular momentum is conserved for one particle moving about a fixed center of force. Thus the angular momentum

$$\mathbf{J} = \mathbf{r} \times \mu\dot{\mathbf{r}} \tag{55}$$

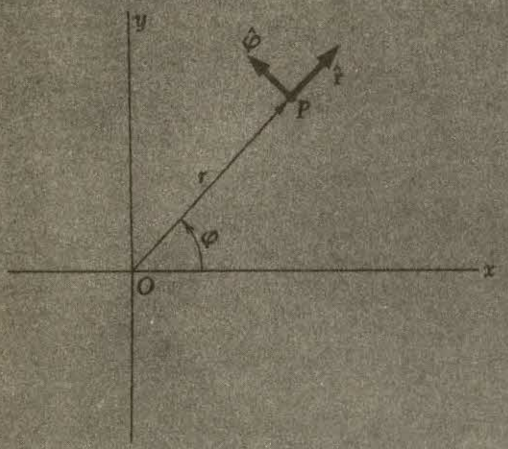
in our effective one-particle problem must be conserved. It is constant in both direction and magnitude. This leads to important simplifications in the solution of (54). It is easy to show that



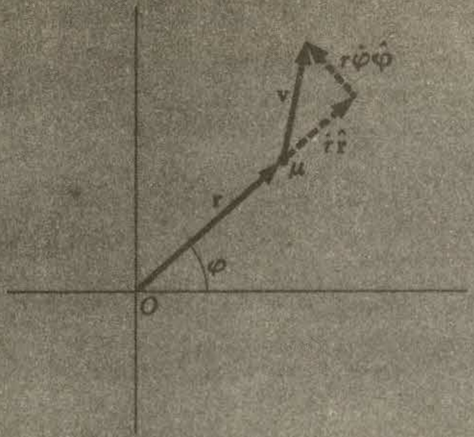
In an inertial frame in which the center of mass is at rest $\mathbf{p}_1 = -\mathbf{p}_2$. The angular momentum of M_1 about the center of mass *plus* the angular momentum of M_2 about the center of mass is a constant, the total angular momentum $\mathbf{J}$.



Since $\mathbf{J}$ is constant, central force motion is confined to a plane.



Plane polar coordinates suitable for describing central force motion of reduced mass μ, at point P, about fixed force center O; $\hat{\mathbf{r}}$ and $\hat{\varphi}$ are unit vectors.



Resolution of the velocity v of the particle into radial and angular components. The kinetic energy is

$$K = \tfrac{1}{2}\mu v^2 = \tfrac{1}{2}\mu(\dot{r}^2 + r^2\dot{\varphi}^2).$$

The total energy is

$$E = K + U = \frac{1}{2}\mu\dot{r}^2 + \frac{\mu r^2}{2}\dot{\varphi}^2 + U.$$

$\mathbf{J}$ in (55) is identical with the usual expression $\mathbf{J} = \mathbf{r}_1 \times M_1\dot{\mathbf{r}}_1 + \mathbf{r}_2 \times M_2\dot{\mathbf{r}}_2$, where $\mathbf{r}_1$, $\mathbf{r}_2$ now are referred to the center of mass $\mathbf{R}_{\text{c.m.}}$ as origin. Simply write out (55) as

$$\mathbf{J} = \left(\frac{M_1M_2}{M_1 + M_2}\right)(\mathbf{r}_1 - \mathbf{r}_2) \times (\dot{\mathbf{r}}_1 - \dot{\mathbf{r}}_2), \tag{55a}$$

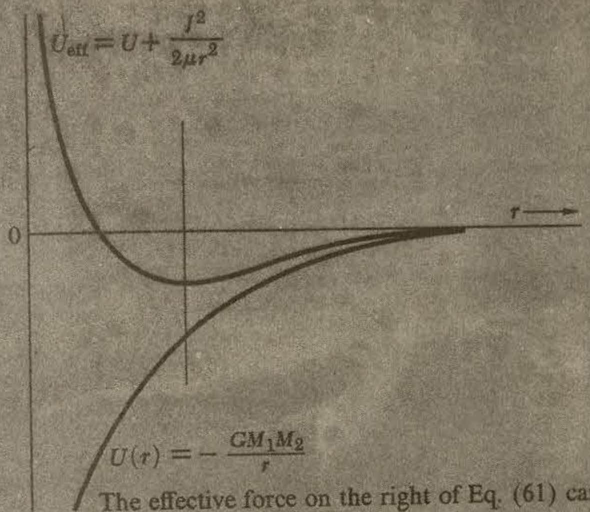
and substitute the relations which follow from (46) with $\dot{\mathbf{R}}_{\text{c.m.}} = 0$; namely, $\dot{\mathbf{r}}_1 = -(M_2/M_1)\dot{\mathbf{r}}_2$ and $\dot{\mathbf{r}}_2 = -(M_1/M_2)\dot{\mathbf{r}}_1$.

We see from (55) that the motion must lie in a plane, as $\mathbf{r} \times \dot{\mathbf{r}}$ *must be constant in direction if* $\mathbf{J}$ *is to be constant in direction.* The position of a particle in a plane need be specified by only two coordinates; we choose $\hat{\mathbf{r}}$ and $\hat{\varphi}$. Here $\hat{\mathbf{r}}$ is the unit vector from a fixed origin and pointing toward the particle; $\hat{\varphi}$ is a unit vector in the plane and normal to $\hat{\mathbf{r}}$. (See the accompanying figure.) The velocity is $\dot{\mathbf{r}}$ and will have a component along $\hat{\mathbf{r}}$ and a component along $\hat{\varphi}$:

$$\dot{\mathbf{r}} = \dot{r}\hat{\mathbf{r}} + r\dot{\varphi}\hat{\varphi} = \dot{r}\hat{\mathbf{r}} + r\omega\hat{\varphi}, \tag{56}$$

where $\dot{\varphi} \equiv \omega$. With this result the angular momentum (55) may be written as

$$\mathbf{J} = \mathbf{r} \times \mu(\dot{r}\hat{\mathbf{r}} + r\omega\hat{\varphi}) = \mu r^2\omega\hat{\mathbf{z}}, \tag{57}$$



The effective force on the right of Eq. (61) can be derived as $-\partial U_{\text{eff}}/\partial r$ of the effective potential

$$U_{\text{eff}}(r) = U(r) + \frac{J^2}{2\mu r^2}.$$

The effective potential is useful in thinking about the effects (or the motion of a particle) of the conservation of angular momentum. It shows us that, when angular momentum is conserved, there is an effective repulsion at small distances.

where $\hat{\mathbf{z}}$ is the unit vector $\hat{\mathbf{r}} \times \hat{\boldsymbol{\varphi}}$ normal to the plane of the motion. We have used the fact that $\mathbf{r} \times \hat{\mathbf{r}} = 0$.

Now consider the equation for the scalar $\ddot{r}$. Recall from Chap. 3 that a rotating coordinate system gives rise to a centrifugal force $\mu\omega^2 r$ which is directed radially outward. The equation of motion in the rotating coordinate system is

$$\mu\ddot{r} = \text{gravitational force} + \text{centrifugal force}, \tag{58}$$

or

$$\mu\ddot{r} = -\frac{GM_1M_2}{r^2} + \mu\omega^2 r. \tag{59}$$

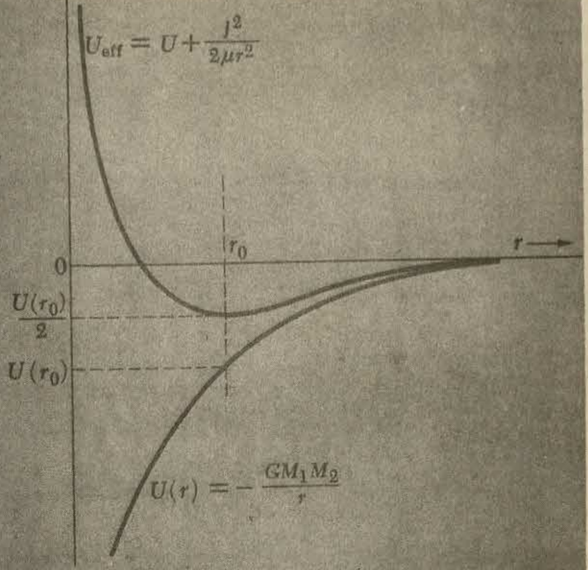
In this equation both r and ω may vary with time.

Now use (57) to express ω in terms of J:

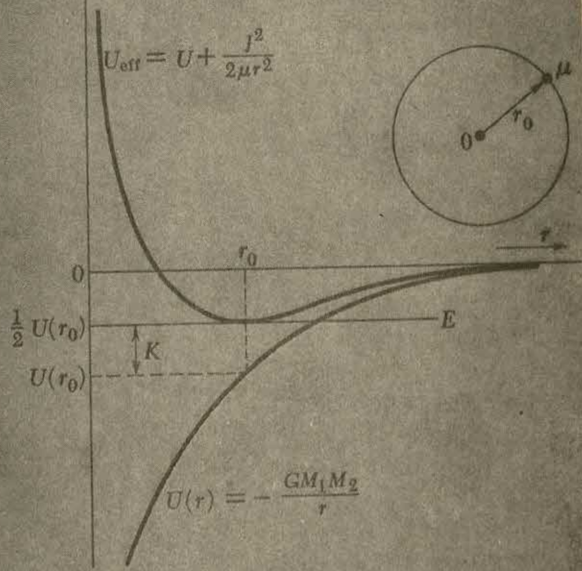
$$\omega^2 = \frac{J^2}{\mu^2 r^4}, \tag{60}$$

so that (59) may be written as

$$\mu\ddot{r} = -\frac{GM_1M_2}{r^2} + \frac{J^2}{\mu r^3}, \tag{61}$$



U_{eff} has a minimum at r_0. At that point $U_{\text{eff}} = \frac{1}{2}U$.



The total energy cannot be less than the minimum of U_{eff}. If it is just that value, $\mu\dot{r}^2/2 = 0$ and $r = r_0$. We have circular motion, with

$$K = E - U = -\frac{U(r_0)}{2} = \frac{GM_1M_2}{2r_0}.$$

where only r depends on the time, because the angular momentum is constant. The solution of (61) is of classical importance in the history of celestial mechanics. It governs the movement of the planets, of double stars, and of satellites.

It is convenient to solve (61) for r as a function $r(\varphi)$ of the angle φ. We have

$$\frac{dr}{dt} = \frac{dr}{d\varphi}\frac{d\varphi}{dt} = \frac{dr}{d\varphi}\omega = \frac{dr}{d\varphi}\cdot\frac{J}{\mu r^2}, \tag{62}$$

using (57). Further, on differentiating once more,

$$\begin{aligned}\frac{d^2r}{dt^2} &= \frac{d^2r}{d\varphi^2}\cdot\left(\frac{J}{\mu r^2}\right)^2 + \frac{dr}{d\varphi}\cdot\frac{J}{\mu}\frac{d}{dt}\left(\frac{1}{r^2}\right)\\ &= \frac{d^2r}{d\varphi^2}\left(\frac{J}{\mu r^2}\right)^2 - \frac{2}{r^3}\cdot\frac{J}{\mu}\cdot\left(\frac{dr}{d\varphi}\right)^2\cdot\frac{J}{\mu r^2},\end{aligned} \tag{63}$$

using (57) again, and also (62).

We consider the function $w(\varphi)$ defined by

$$w(\varphi) = \frac{1}{r(\varphi)}, \tag{64}$$

because this function has been found (after a hundred years of experience) to lead to a much simpler differential equation than does r. Let us differentiate (64) to form

$$\frac{dw}{d\varphi} = -\frac{1}{r^2}\frac{dr}{d\varphi}; \qquad \frac{d^2w}{d\varphi^2} = -\frac{1}{r^2}\frac{d^2r}{d\varphi^2} + \frac{2}{r^3}\left(\frac{dr}{d\varphi}\right)^2. \tag{65}$$

On comparison of (65) with (63) we see that

$$\frac{d^2r}{dt^2} = -\frac{1}{r^2}\left(\frac{J}{\mu}\right)^2\frac{d^2w}{d\varphi^2}, \tag{66}$$

so that the equation of motion (61) may be written as

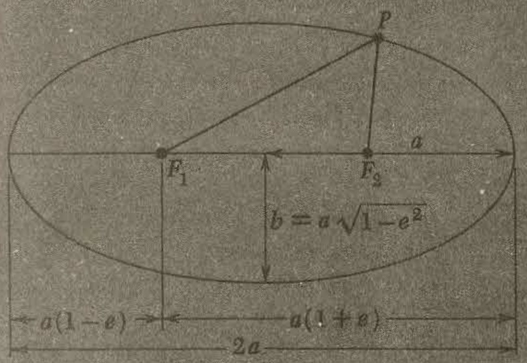
$$-w^2\frac{J^2}{\mu}\frac{d^2w}{d\varphi^2} = -w^2GM_1M_2 + w^3\frac{J^2}{\mu}, \tag{67}$$

or

$$\boxed{\frac{d^2w}{d\varphi^2} + w = \frac{\mu GM_1M_2}{J^2}.} \tag{68}$$

Notice that the right-hand side is a constant, so that we have a very simple equation.

The left-hand side of (68) is reminiscent of the equation of motion of the harmonic oscillator. We see that for any constant



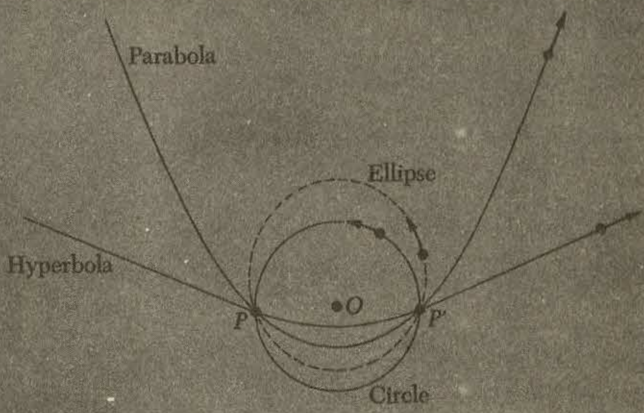
Properties of the ellipse: For any point P the distance $F_1P + F_2P = \text{const} = 2a$. The equation of the ellipse is

$$r = \frac{a(1-e^2)}{(1 - e\cos\varphi)}, \qquad 0 < e < 1.$$

The semiminor axis is given by

$$b = a\sqrt{1-e^2}.$$

The area of the ellipse is πab.



Orbits of particles with the same reduced mass μ and angular momentum **J**, but different energies E, about fixed center of force O. The orbits all cross at P and P'.

Orbit	*Eccentricity*
Circle	$e = 0$ } $E < 0$
Ellipse	$e = \frac{1}{3}$ } $E < 0$
Parabola	$e = 1, E = 0$
Hyperbola	$e = 3, E > 0$

A the quantity $w = A\cos\varphi$ is a solution of

$$\frac{d^2w}{d\varphi^2} + w = 0. \tag{69}$$

If to w we add the constant $\mu GM_1M_2/J^2$, we have

$$w = A\cos\varphi + \frac{\mu GM_1M_2}{J^2}, \tag{70}$$

which is a solution of (68). Try it! Recall that $w \equiv 1/r$.

Now (70) is just what is called the polar form of the equation of a conic section (ellipse, circle, parabola, or hyperbola). You may recall from your favorite text on analytic geometry that the equation of a general conic section (section of a cone by a plane) may be written as

$$\frac{1}{r} = \frac{1}{se}(1 - e\cos\varphi). \tag{71}$$

The constant e is known as the *eccentricity*. The constant s determines the scale of the figure. The four types of possible curves described by (71) are

Hyperbola	$e > 1$,
Parabola	$e = 1$,
Ellipse	$0 < e < 1$,
Circle	$e = 0$.

A crude but effective way to satisfy yourself that this equation can give a curve which at least looks like an ellipse is to calculate r for a range of values of φ. The results may be plotted conveniently on polar graph paper, which is readily available. Such graph paper is marked out with lines at constant radius and constant angle. We tabulate below our rough calculations from (71)

φ	$\cos\varphi$	$2(1 - \frac{1}{2}\cos\varphi)$	r
0°	1.00	1.00	1.00
20°	0.94	1.06	0.94
40°	0.77	1.23	0.81
60°	0.50	1.50	0.67
80°	0.17	1.83	0.55
90°	0.00	2.00	0.50
100°	−0.17	2.17	0.46
120°	−0.50	2.50	0.40
140°	−0.77	2.77	0.36
160°	−0.94	2.94	0.34
180°	−1.00	3.00	0.33

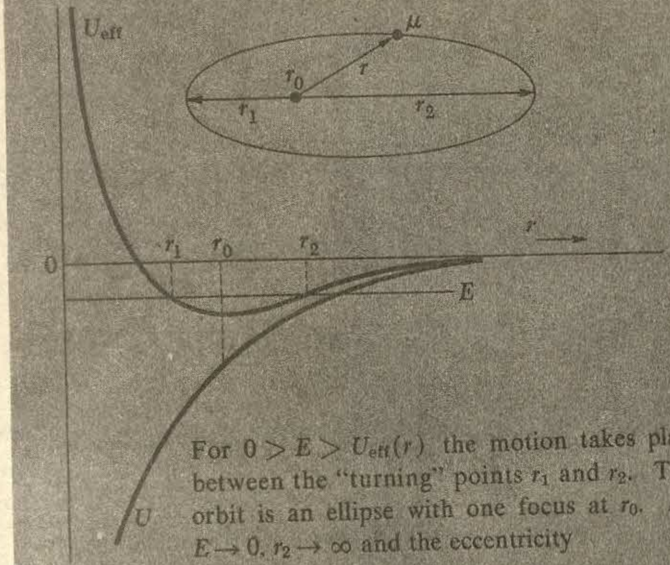
for the case $s = 1$ and $e = \frac{1}{2}$; you should plot these values of r vs φ on polar graph paper and confirm that the curve looks like an ellipse. Make similar calculations for $s = 1$, $e = 2$; this curve is a hyperbola.

By comparison of (70) with (71) we find

$$\frac{1}{se} = \frac{\mu G M_1 M_2}{J^2}; \qquad \frac{1}{s} = -A. \tag{72}$$

We now wish to express e (or s) in terms of the total energy E of the system. (The energy is another constant of the motion.) To do this most easily we note that the energy takes on a particularly simple form at the minimum and maximum distances of the particle from the origin, where the velocity vector is perpendicular to $\mathbf{r}$. The kinetic energy at these points is of a simple form (because $\dot{r} = 0$ at these points):

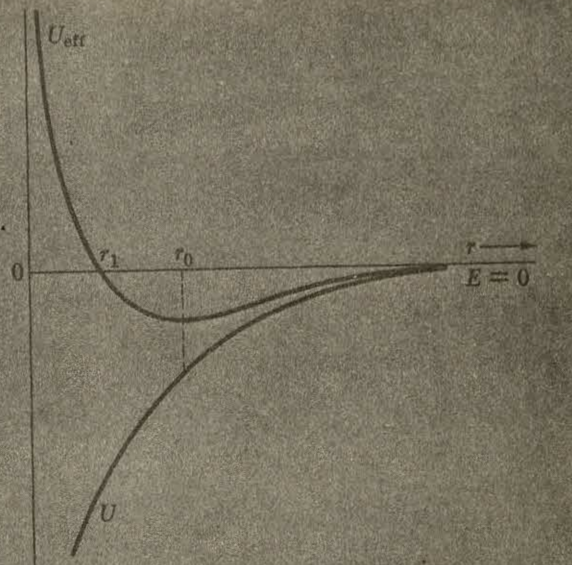
$$K = \frac{1}{2}\mu\dot{\mathbf{r}}\cdot\dot{\mathbf{r}} = \frac{1}{2}\mu r^2\dot{\varphi}^2 = \frac{1}{2}\mu\left(\frac{J}{\mu r}\right)^2, \tag{73}$$



For $0 > E > U_{\text{eff}}(r)$ the motion takes place between the "turning" points r_1 and r_2. The orbit is an ellipse with one focus at r_0. As $E \to 0$, $r_2 \to \infty$ and the eccentricity

$$e = \frac{r_2 - r_1}{r_1 + r_2} \to 1,$$

as for a parabola.



For $E = 0$, $r_2 = \infty$ and $e = 1$. The orbit is a parabola.

using (56) and (57). (At all other points of the orbit the kinetic energy contains a contribution from $\dot{r}^2$.) From (73) we have at r_{min} and r_{max} the relation

$$E = K + U = \frac{J^2}{2\mu}\left(\frac{1}{r_{\text{min}}}\right)^2 - \frac{GM_1M_2}{r_{\text{min}}}$$
$$= \frac{J^2}{2\mu}\left(\frac{1}{r_{\text{max}}}\right)^2 - \frac{GM_1M_2}{r_{\text{max}}}. \quad (74)$$

Now from (71) we see that, with $\varphi = \pi$ and $\varphi = 0$,

$$\frac{1}{r_{\text{min}}} = \frac{1}{se}(1 + e); \qquad \frac{1}{r_{\text{max}}} = \frac{1}{se}(1 - e). \quad (75)$$

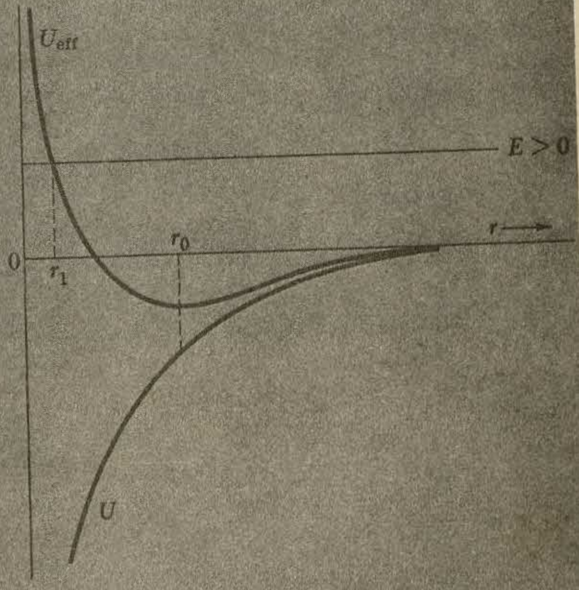
On substituting (75) into (74) and using (72), after some algebra we find the result

$$e = \left(1 + \frac{2EJ^2}{\mu G^2 M_1{}^2 M_2{}^2}\right)^{1/2}; \qquad E = -\frac{\mu G M_1{}^2 M_2{}^2}{2J^2}(1 - e^2). \quad (76)$$

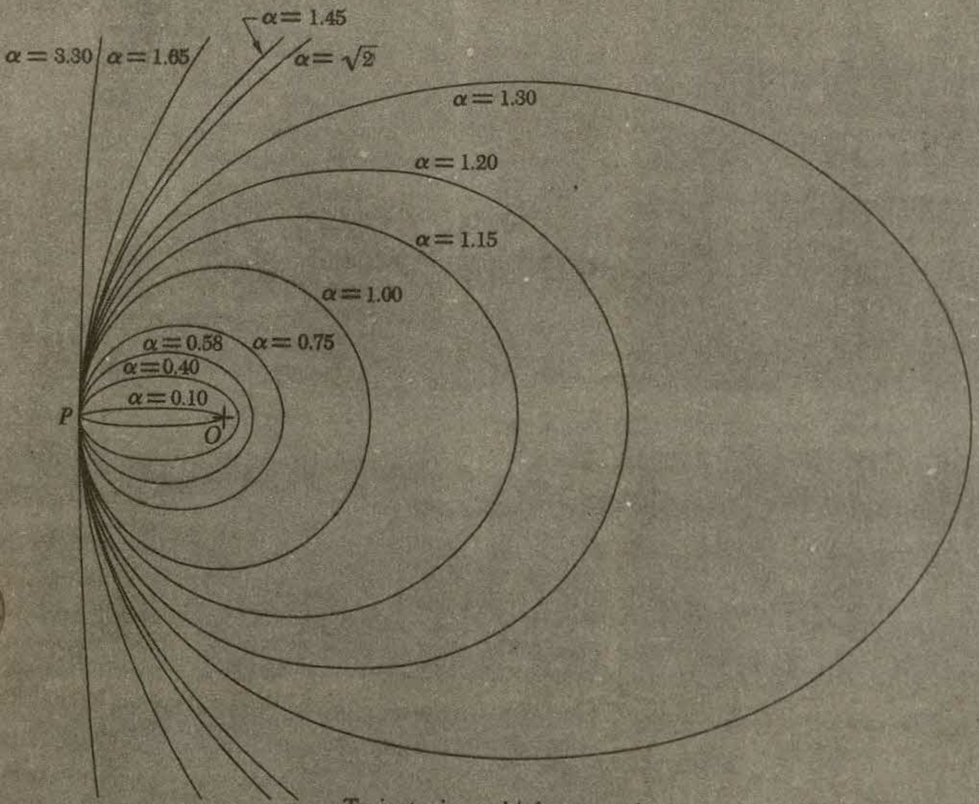
Note from (75) that

$$\frac{r_{\text{max}} - r_{\text{min}}}{r_{\text{max}} + r_{\text{min}}} = e. \quad (77)$$

This is a convenient relation to remember for the eccentricity of an ellipse.



For $E > 0$, $e > 1$ and the orbit is hyperbolic. The distance r_1 is called the distance of closest approach.



Trajectories which pass through a common point P and normal to the line OP from the force center at O. With $\mathbf{v}_0$ as the velocity of the circular orbit, the parameter α is defined by

$$v_P(\alpha) = \alpha v_0.$$

It is shown in Eq. (83) that

$$E(\alpha) = (2 - \alpha^2)E_0.$$

Example. Circular orbit. What is the connection between the total energy and the angular momentum in a circular orbit?

For a circular orbit $e = 0$, so that from (76) we have

$$E = -\frac{\mu G^2 M_1{}^2 M_2{}^2}{2J^2}. \tag{78}$$

The total energy is negative referred to $V(\infty) = 0$ because the system is bound.

We find that some students are inclined to think all closed orbits should be circular. To gain a feeling for elliptical orbits, study the accompanying figure. In the figure we see a family of trajectories of a particle attracted by an inverse-square force toward the origin at O, as denoted by the cross. The family has been chosen so that all the trajectories pass through a common point P, and at P the velocity is required to be perpendicular to the line between O and P. The different orbits are characterized by different values of the velocity at P. The general velocity v_P is conveniently written as

$$\frac{v_P}{v_O} \equiv \alpha, \tag{79}$$

where v_O is the velocity of the *circular* orbit centered at O and passing through P. For $\alpha = 1$, the orbit is circular; for $\alpha < \sqrt{2}$, the orbit is an ellipse; for $\alpha = \sqrt{2}$, the orbit is a parabola; and for $\alpha > \sqrt{2}$, the orbit is a hyperbola. [These results are illuminated by (83) below.]

By calculation of the energy, we can verify that the transition between open and closed trajectories occurs for $\alpha = \sqrt{2}$. At the point P the total energy may be written as

$$E = \frac{1}{2}\mu v_P{}^2 - \frac{B}{r_O} = \frac{1}{2}\mu\alpha^2 v_O{}^2 - \frac{B}{r_O}$$

$$= \frac{1}{2}(\alpha^2 - 1)\mu v_O{}^2 + \frac{1}{2}\mu v_O{}^2 - \frac{B}{r_O} = E_O + \frac{1}{2}(\alpha^2 - 1)\mu v_O{}^2, \tag{80}$$

where E_O and v_O refer to the energy and velocity of the circular orbit; B is a constant equal to GM_1M_2; r_O is the distance between P and O. Now in the circular orbit

$$\frac{\mu v_O{}^2}{r_O} = \frac{B}{r_O{}^2}, \tag{81}$$

where the term on the left is the mass times the centripetal acceleration, and the term on the right is the gravitational force. Using this result, we can write the energy in the circular orbit as

$$E_O = \frac{1}{2}\mu v_O{}^2 - \frac{B}{r_O} = \frac{1}{2}\mu v_O{}^2 - \mu v_O{}^2 = -\frac{1}{2}\mu v_O{}^2, \tag{82}$$

and (80) may be written

$$E = E_O - (\alpha^2 - 1)E_O = (2 - \alpha^2)E_O = (\alpha^2 - 2)|E_O|. \tag{83}$$

If $\alpha^2 > 2$, the total energy is positive and the orbit is open. If $\alpha^2 < 2$, the total energy is negative and the orbit is closed: the particle cannot escape to an infinite distance.

Kepler's Laws. Kepler's determination that the orbits of the planets are ellipses about the sun was one of the greatest experimental discoveries in the history of mankind. Together with his formulation of the empirical laws of planetary motion it provided the original experimental evidence for Newton's laws of mechanics and for the theory of gravitational attraction. Kepler stated the three laws essentially as follows:

I. All planets move in an elliptical path, with the sun at one focus.

II. A line drawn from the sun to a planet sweeps out equal areas in equal times.

III. The squares of the periods of revolution of the several planets about the sun are proportional to the cubes of the semimajor axes of the ellipses. (This statement is more general than the original formulation by Kepler.) In our entire discussion we neglect the effects of the other planets on the one under consideration.

We have demonstrated just above that closed orbits are elliptical. Kepler's second law was demonstrated in Chap. 6, Eq. (65), where it was shown to be simply a statement of the conservation of angular momentum.

We are now going to derive Kepler's third law. If dS is the area swept out in time dt by the radius vector from the sun to the planet, then we found that

$$\frac{dS}{dt} = \frac{J}{2\mu} = \text{const}, \tag{84}$$

where J is the angular momentum and μ is the reduced mass. On integrating (84) over one period T of the motion we have

$$S = \frac{JT}{2\mu}; \qquad \text{or} \qquad T = \frac{2S\mu}{J} = \frac{2\pi ab\mu}{J}. \tag{85}$$

Here $S = \pi ab$ is the area of an ellipse with a as the semimajor axis and b as the semiminor axis.

Now $2a = r_{\max} + r_{\min}$ is an obvious property of an ellipse; with the use of (75) we have

$$2a = \frac{se}{1+e} + \frac{se}{1-e} = \frac{2se}{1-e^2}. \tag{86}$$

With (72) this becomes

$$2a = \frac{2}{1 - e^2} \cdot \frac{J^2}{GM_1M_2\mu}. \qquad (87)$$

On squaring (85) and using (87) for J^2, we have

$$T^2 = \frac{(2\pi ab\mu)^2}{aGM_1M_2\mu(1 - e^2)} = \frac{4\pi^2 ab^2\mu}{GM_1M_2(1 - e^2)}. \qquad (88)$$

But it is a property of the eccentricity e that

$$b^2 = a^2(1 - e^2), \qquad (89)$$

whence (88) reduces to

$$\boxed{T^2 = \frac{4\pi^2 a^3}{G(M_1 + M_2)}.} \qquad (90)$$

You should verify (90) for a circular orbit; this can be done easily, because in a circular orbit the balance of forces gives us $\mu\omega^2 r = GM_1M_2/r^2$, so that $\omega^2 r^3$ is constant.

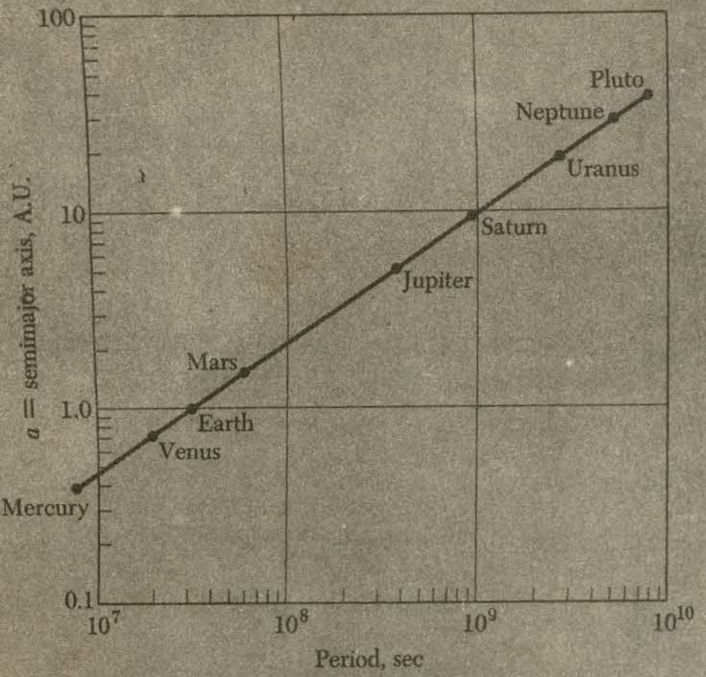
The table gives details of the orbits of the major planets. Notice that the orbit of Earth is very nearly circular. An *astronomical unit* (A.U.) of length is defined as one-half of the sum of the longest and the shortest distances of Earth from the sun.

$$1 \text{ A.U.} = 1.495 \times 10^{13} \text{ cm}.$$

This unit is not to be confused with the parsec. A *parsec* is the distance at which one astronomical unit subtends an angle of one second of arc.

$$1 \text{ parsec} = 3.084 \times 10^{18} \text{ cm}.$$

	Semimajor axis, A.U.	*Period, sec*	*Eccentricity*	*Inclination*	*Mass, gm*
Mercury	0.387	7.60×10^6	0.205	7°00	3.28×10^{26}
Venus	0.723	1.94×10^7	0.006	3°23′	4.83×10^{27}
Earth	1.000	3.16×10^7	0.016	....	5.98×10^{27}
Mars	1.523	5.94×10^7	0.093	1°51′	6.37×10^{26}
Jupiter	5.202	3.74×10^8	0.048	1°18′	1.90×10^{30}
Saturn	9.554	9.30×10^8	0.055	2°29′	5.67×10^{29}
Uranus	19.218	2.66×10^9	0.046	0°46′	8.80×10^{28}
Neptune	30.109	5.20×10^9	0.008	1°46′	1.03×10^{29}
Pluto	39.60	7.82×10^9	0.246	17° 7′	5.4×10^{27}



From the slope of this straight line we can easily see that the period T varies as $a^{3/2}$.

The distance to the nearest star from the sun is 1.31 parsecs. The inclination as given in the table is the angle between the plane of a planet's orbit and the plane of Earth's orbit (the ecliptic).

Let us test Kepler's third law for the orbit of Uranus compared with that of Earth. The cube of the ratio of the lengths of the semimajor axis is

$$\left(\frac{19.22}{1}\right)^3 \cong 71.0 \times 10^2. \tag{91}$$

The square of ratio of the periods is

$$(84.2)^2 \cong 70.9 \times 10^2, \tag{92}$$

which agrees quite closely with (91). (A 10-in. slide rule was used for the calculation; the student should perform the same calculation for the orbit of Mercury compared with that of Earth.) On the accompanying graph we have used log-log paper to plot the data for the planets. On log-log paper a power law will appear as a straight line; the slope of the line will give the exponent of the power law. (Prove this.)

Newton also tested Kepler's third law against the observed periods of revolution of the four largest moons of Jupiter and found very good agreement.

Further reading

PSSC, chaps. 4 (paragraph 3), 22 (paragraphs 4–8), 25 (paragraph 5), 26 (paragraph 2), and 28 (paragraphs 1–2).

P. van de Kamp, *Elements of astromechanics* (W. H. Freeman, San Francisco, 1964). Paperback; selected elementary topics in celestial mechanics.

O. Struve, B. Lynds, and H. Pillans, *Elementary astronomy* (Oxford University Press, New York, 1959). Emphasizes the main ideas of physics in relation to the universe; excellent.

T. S. Kuhn, *The Copernican revolution* (Vintage paperback, Division of Random House, New York, 1962).

American Association of Physics Teachers, Selected Reprints—*Kinematics and Dynamics of Satellite Orbits* (American Institute of Physics, New York, 1963).

Problems

1. *Reduced mass.* The fundamental vibration frequency of a diatomic molecule may be written

$$\omega = \left(\frac{C}{\mu}\right)^{1/2},$$

where C is the force constant analogous to a macroscopic spring, and μ is the reduced mass for the two atoms. The vibrational frequency of the carbon monoxide molecule CO is $\omega \simeq 0.6 \times 10^{15}\ \text{sec}^{-1}$.

(*a*) Calculate μ in grams from the atomic masses. *Ans.* 1×10^{-23} gm.

(*b*) Calculate the force constant C for stretching in this molecule. *Ans.* 4×10^{6} dyne/cm.

2. *Gravitational attraction of infinite line.* Show that $2G\rho M_1/R$ is the gravitational force on a mass M_1 located at a distance R from an infinite line of mass density ρ per unit length. (Be careful about the direction of the force due to a line element.)

3. *Gravitational attraction of finite line.* You are at a point x on the perpendicular bisector of a line of length $2L$. The mass of the line is M; the origin of the coordinate system is on the line.

(*a*) Find an expression for the potential energy, referred to $U = 0$ at $x = \infty$. *Ans.* $-(GMm/L) \log \{[L + (x^2 + L^2)^{1/2}]/x\}$.

(*b*) Find an expression for the gravitational force exerted by the line on a point mass m at x. In what direction is the force?

(*c*) Show that your result for part (*a*) reduces to $U \cong -GMm/x$ when $x \gg L$.

Consider a thin wire of length 2 m and linear density 2 gm/cm.

(*d*) What is the value (in dynes) of the gravitational force exerted by the wire on a point mass $M = 0.5$ gm located on the long axis of the wire at a distance of 3 m from the center of the wire? *Ans.* 1.7×10^{-10} dyne.

(*e*) What is the potential energy (in ergs) of the point mass in the force field of the wire, at the position given in part (*a*)? *Ans.* -4.6×10^{-7} erg.

4. *Inverse-2.1-power force.* Find the potential energy of a test particle of mass M_1 located *inside* a spherical shell of radius R at a distance r from its center. The mass of the shell is $4\pi\sigma R^2$. The force law between two point particles of masses M_1, M_2 is assumed to be

$$F = -G' \frac{M_1 M_2}{r^{2.1}};$$

here G' is a constant. No such force law is known to physics, but note the sensitivity of the result to the deviation of the exponent of r from the value 2.0.

5. *Gravitational potential energy of array of stars.* Find the mutual gravitational potential energy (in ergs) of a system of eight stars, each of mass equal to that of the sun, located at the corners of a cube whose edge is 1 parsec. (Omit the self-energy of each star.) *Ans.* 2×10^{42} gm-cm^2/sec^2.

6. *Gravitational acceleration at surface of white dwarf star.* Consider a white dwarf star with mass equal to that of the sun and radius equal to 0.02 that of the sun. These data apply roughly to the best-known white dwarf, Sirius B, which is the companion of the brightest star on the celestial sphere, Sirius. White dwarfs are believed to be the end products of one phase of stellar evolution.

(*a*) What is the value of g, the acceleration due to gravity, at the surface of Sirius B?

(*b*) What is the density of Sirius B? *Ans.* 2×10^5 gm/cm^3.

7. *Orbital motion of binary stars.* The most massive star known at present is J. S. Plaskett's star. It is a double, or binary, star†; that is, it consists of two stars bound together by gravity. From spectroscopic studies it is known that

(*a*) The period of revolution about their center of mass is 14.4 days (1.2×10^6 sec).

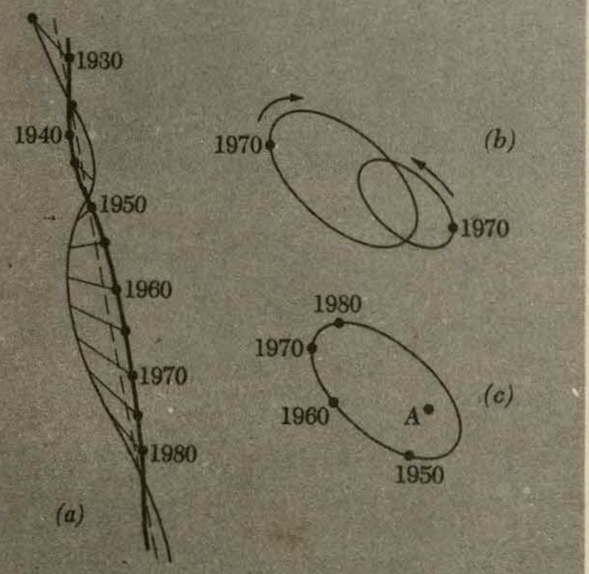
(*b*) The velocity of each component is about 220 km/sec. Because both components have nearly equal (but opposite) velocities, we may infer that they are nearly equidistant from the center of mass, and hence that their masses are nearly equal.

(*c*) The orbit is nearly circular.

From these data calculate the reduced mass and the separation of the two components. *Ans.:* $\mu \cong 0.6 \times 10^{35}$ gm; separation $\cong 0.8 \times 10^{13}$ cm.

8. *Motion in a galaxy.* Consider a uniform spherical distribution of stars in a galaxy of total mass M and radius R_0. A star of mass M_s at some distance $r < R_0$ from the center will move under the action of a central force whose magnitude depends on the mass included within a sphere of radius r.

(*a*) What is the force at r?



The visual orbit of Sirius. The heavy curve in (*a*) shows the sinusoidal motion of the primary star, the thin curve represents the sinusoidal motion of the white dwarf companion, the dashed curve is the motion of the center of gravity of the system; (*b*) shows the apparent orbits of the two components around their common center of gravity; and (*c*) is the apparent orbit of the companion about the primary. (*After Struve, Lynds, and Pillans*)

---

† There is a good discussion of binary stars in O. Struve, B. Lynds, and H. Pillans, *Elementary astronomy*, chap. 29 (Oxford University Press, New York, 1959). At least half of the fifty stars nearest to the sun are binary or multiple stars.

(*b*) What is the circumferential velocity of the star if it moves about the center in a circular orbit?

9. *Elliptic orbit relations.* (*a*) Show that the angular momentum of a planet in an elliptic orbit through a point P, as defined in the text above Eq. (79), is given by $J = \alpha J_0$, where J_0 refers to a circular orbit through P, and α is defined by Eq. (79).

(*b*) Show that the eccentricity is

$$e = |\alpha^2 - 1|.$$

10. *Motion of a dwarf galaxy.* It has been found recently that our galaxy is surrounded by several (at least six) very small dwarf galaxies. Their proximity, low mass, and low relative velocity with respect to our galaxy (those which have been measured have velocities less than 10^7 cm/sec) suggest that they are gravitationally bound to our stellar system. Consider one of these, the so-called Sculptor system. From measurements on certain of its variable stars, its distance from the center of our galaxy is found to be 2×10^{23} cm. The total mass of the Sculptor galaxy is estimated from its luminosity to be 3×10^6 times that of the sun. The mass of our galaxy is estimated to be approximately 4×10^{11} times that of the sun. Assume that the Sculptor system revolves around our galaxy in a circular orbit.

(*a*) Calculate the period of revolution.

(*b*) Find the orbital velocity.

(*c*) Calculate the relative acceleration of the dwarf galaxy.

11. *Meteoroid orbits.* Meteoroids are small bodies in interplanetary space, moving in closed orbits about the sun. Occasionally they collide with the earth's atmosphere, producing visible meteors, which are observed to have velocities ranging from 1.1×10^6 to 7.5×10^6 cm/sec at altitudes up to 10^7 cm above the earth's surface. The orbit will be closed as long as the velocity at any point is less than the escape velocity for that distance from the sun:

$$v_{\max} = \left(\frac{2GM_s}{R}\right)^{1/2}.$$

We neglect friction, and also the attraction of the earth on the meteoroid. What is the value of $v_{\max}$ appropriate to a point on the earth's orbit?

12. *Gravitational pressure at center of earth.* Consider a homogeneous spherical mass in hydrostatic equilibrium. The radius is R; and the mass density is ρ.

(*a*) Show that the pressure at a distance r from the center is

$$p = \frac{2\pi}{3}\rho^2 G(R^2 - r^2).$$

(*b*) For the earth, take a radius $R = 6.3 \times 10^8$ cm and a uniform density $\rho = 5.5$ gm/cm^3, and calculate the pressure at the center.

13. *Measurement of G.* Very sensitive measurements can be made with quartz fibers. The tensile strength of a fiber varies as the square of the radius, while the torsion constant varies as the fourth power of the radius. It is therefore desirable to use the smaller radii, if we wish the high sensitivity that is achieved with low values of the torsion constant. The torsion constant K is defined as the torque per radian, so that $N = -K\varphi$, where N is the torque. It is not unusual to work with fibers having K in the range 0.01 to 1 dyne-cm/rad. Mirrors and electronic systems make it possible, under exceptional conditions, to detect rotations down to 10^{-8} rad. Giving the relevant orders of magnitude of all quantities involved, design an experimental apparatus for measuring the gravitational constant G in an undergraduate laboratory. (Do not expect to detect 10^{-8} rad!) The value of the torsion constant is of the order of

$$K \approx \frac{10^{11}R^4}{L} \text{ dyne-cm/rad,}$$

where R is the radius of the fiber and L the length, in centimeters.

14. *Stellar energy sources.* The total energy generation of the sun, 4×10^{33} ergs/sec, is known from the rate of energy radiation. Suppose the sun has been radiating at this rate for Y years, since the original condensation began. Half the gravitational potential energy in the sun has gone into kinetic energy (according to the virial theorem) and the other half into radiation. Show that $Y \approx 3 \times 10^7$ years. The result for Y is much too short to account for the known age of the sun, 5×10^9 years. (The sun is supposed to be at least as old as the earth.) Nuclear, rather than gravitational, energy is a much better source of the sun's luminosity.

15. *Planck's length.* Show that a quantity having the dimensions of length can be formed from the gravitational constant G, the velocity of light c, and Planck's constant h (which is an energy times a time). What is the order of magnitude of this length? It is sometimes called Planck's length; its importance is still somewhat speculative.

16. *Globular clusters.* A globular cluster of stars is a spherical distribution of perhaps 10^5 stars; the linear diameter of the cluster may be of the order of 40 parsecs.

(*a*) If the stars are uniformly distributed over the volume of the cluster, what is the average number of stars in a cubic parsec?

(*b*) What is the order of magnitude of the gravitational self-energy of a globular cluster? Assume the average stellar mass to be equal to that of the sun. (In computing the gravitational self-energy of the cluster exclude the gravitational self-energy of the individual stars.)

17. *Form of the sea on a uniform earth.* A uniform spherical earth is covered by water of density ρ. The surface of the sea assumes the form of an oblate spheroid (a flattened sphere) when the earth spins with angular velocity ω. Find an approximate expression for the difference in the depth of the sea at a pole and at the equator, assuming that the surface of the sea is a surface of constant potential energy. (Why is this a plausible assumption?) Neglect the gravitational attraction of the sea upon itself.

The globular cluster M3. (*Lick Observatory photograph*)

Hint: At equal distances r the gravitational force on a unit mass of water is equal at the pole and at the equator, but at the equator there is the additional centrifugal force $\omega^2 r$. Thus we have for the forces

$$F_{\text{eq}}(r) = F_{\text{pole}}(r) + \omega^2 r,$$

and the potential energy of a unit mass on the rotating earth (referred to the center of the earth as the zero of potential energy) is

$$U_{\text{eq}}(r) = U_{\text{pole}}(r) + \frac{1}{2}\omega^2 r^2.$$

Let D_{eq} and D_{pole} denote the deformations from spherical shape. If U_{eq} at $R_E + D_{\text{eq}}$ is to equal U_{pole} at $R_E + D_{\text{pole}}$, show that

$$g(D_{\text{eq}} - D_{\text{pole}}) \cong \frac{1}{2}\omega^2 R_E^2,$$

where g has the usual meaning. Thus

$$\frac{D_{\text{eq}} - D_{\text{pole}}}{R_E} \cong \frac{\omega^2 R_E}{g} \cong \frac{3.4}{980} \cong \frac{1}{575}.$$

This is very close to the value 1/298.4 observed for the actual figure of the earth.

18. *Origin of the tides.* Discuss the origin of the tides. Note that the chief tide-raising force is the difference between the moon's attraction on the surface waters of the earth and its attraction on the earth. For a more detailed explanation you may wish to consult a good encyclopedia or an elementary text on astronomy. Include in your discussion a careful answer to the following question:

Why do simultaneous high tides tend to appear at that point on the earth closest to the moon and also on the opposite side of the earth? That is, why do tides tend to occur twice (and not once) a day.

Advanced Topic 1: Virial theorem

Earlier we stated without proof an important result: *Stable static equilibrium is impossible for particles interacting with inverse-square-law forces.* This means that in a system where inverse-square-law forces are operating, no state in which the kinetic energy of a system vanishes can be stable. This result applies to the electrons in an atom, the atoms in a star, and the stars in a galaxy. The universe at all levels of observation must therefore be composed of moving bodies, and only in some average sense can we consider any part of the natural world to be motionless. We can have a *steady state* in which, in some average sense, no great changes occur, but not a *static state* in which there is no motion.

Use of the *virial theorem*,† gives us the average (over long times) of the kinetic energy of particles bound by inverse-square-law forces:

$$\langle \text{Kinetic energy} \rangle = -\frac{1}{2}\langle \text{potential energy} \rangle, \tag{93}$$

where the angle brackets denote the average over very long times. (It will usually be clear what we mean by "very long" in any specific application.) The potential energy as used here is defined to be zero at infinite separation of all the particles.

We now prove the virial theorem for inverse-square-law forces. First consider the motion of a single particle in the central field described by the potential energy

$$U(r) = \frac{C}{r}, \tag{94}$$

---

† The virial theorem is not restricted to inverse-square-law forces; we discuss it here only for such forces because of their great importance in physics. The factor $-\frac{1}{2}$ that appears in (93), however, holds only for inverse-square-law forces.

where C is a constant. The equation of motion is

$$\mathbf{F} = \frac{C}{r^2}\hat{\mathbf{r}} = M\dot{\mathbf{v}}, \tag{95}$$

for velocities in the nonrelativistic regime. Take the scalar product with $\mathbf{r}$ of the terms in (95):

$$\mathbf{r}\cdot\mathbf{F} = M\mathbf{r}\cdot\dot{\mathbf{v}} = \frac{C}{r} = U(r). \tag{96}$$

Note further that

$$M\frac{d}{dt}(\mathbf{r}\cdot\mathbf{v}) = M\dot{\mathbf{r}}\cdot\mathbf{v} + M\mathbf{r}\cdot\dot{\mathbf{v}} = M\mathbf{v}\cdot\mathbf{v} + M\mathbf{r}\cdot\dot{\mathbf{v}}. \tag{97}$$

We combine (96) and (97) to obtain

$$M\frac{d}{dt}(\mathbf{r}\cdot\mathbf{v}) = 2K + U; \tag{98}$$

here $K = \frac{1}{2}Mv^2$ is the kinetic energy. Thus

$$\frac{1}{2}M\frac{d}{dt}(\mathbf{r}\cdot\mathbf{v}) = \frac{1}{2}Mv^2 + \frac{C}{2r} = \frac{1}{2}Mv^2 + \frac{1}{2}U(r), \tag{99}$$

where $\mathbf{r}$ and $\mathbf{v}$ vary with time. For an attractive potential C is negative, and it is possible to have a bound state in which the particle remains indefinitely within a finite volume around the force center. For a bound state the particle sooner or later must reverse its direction of motion and $\mathbf{r}\cdot\mathbf{v}$ must have an upper bound. This quantity will increase as often as it decreases, on the average. The time average of $d(\mathbf{r}\cdot\mathbf{v})/dt$ in a bound state must be zero if we average over many cycles of the motion. Thus for a bound state the time average of (99) gives:

$$\tfrac{1}{2}M\langle v^2\rangle = -\tfrac{1}{2}\langle U\rangle, \tag{100}$$

or

> The average kinetic energy equals minus one-half the average potential energy, for a particle bound by an inverse-square-law attraction.

Eq. (100) is called the *virial theorem*. The theorem does *not* mean that at any instant the kinetic energy and potential energy are related by (100); the theorem is a statement about averages over long periods of time. (On your first reading of this chapter you may wish to skip ahead from this point to the examples.)

The same theorem also holds for any number of particles kept within some finite volume by mutual inverse-square-law attractive forces, even if all masses are not equal and even if some of the forces are repulsive (for example, a mixture of electrons and nuclei forming a molecule). To prove the result, consider N particles (designated 1, 2, 3, . . . , N) of masses M_1, $M_2, \ldots, M_N$. We write the potential energy between the ith and jth particles as

$$U_{ij} = \frac{C_{ij}}{r_{ij}}, \tag{101}$$

where C_{ij} is a positive or negative constant and $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$. By definition

$$C_{ji} = C_{ij}. \tag{102}$$

The total potential energy is

$$U = \frac{1}{2}\sum_{i=1}^{N}{}' \sum_{j=1}^{N} \frac{C_{ij}}{r_{ij}}, \tag{103}$$

where the fraction $\frac{1}{2}$ enters because in the double sum each pair is counted twice. The prime on one of the Σ's means that we exclude the terms $i = j$ from the sum. We now prove the generalized virial theorem by beginning with the N equations of motion, one for each particle:

$$M_i\dot{\mathbf{v}}_i = \sum_{j=1}^{N}{}' \frac{C_{ij}}{r_{ij}^3}(\mathbf{r}_i - \mathbf{r}_j). \tag{104}$$

We first take the dot product of $\mathbf{r}_i$ with both sides and sum over all particles:

$$\sum_{i=1}^{N} M_i(\mathbf{r}_i \cdot \dot{\mathbf{v}}_i) = \sum_{i=1}^{N}{}' \sum_{j=1}^{N} \frac{C_{ij}}{r_{ij}^3}(\mathbf{r}_i - \mathbf{r}_j) \cdot \mathbf{r}_i. \tag{105}$$

Each term on the left-hand side can be reexpressed in the manner of (97), so that the left-hand side, $\mathfrak{L}$, of (105) becomes:

$$\mathfrak{L} = -\sum_{i=1}^{N} M_i v_i^2 + \frac{d}{dt}\sum_{i=1}^{N} M_i(\mathbf{r}_i \cdot \mathbf{v}_i); \tag{106}$$

or

$$\mathfrak{L} = -2(\text{kinetic energy}) + \frac{1}{2}\frac{d^2}{dt^2}\sum_{i=1}^{N} M_i \mathbf{r}_i^2. \tag{106a}$$

In this last form we used the identity

$$\frac{d}{dt}\mathbf{r} \cdot \mathbf{r} \equiv 2\mathbf{r} \cdot \dot{\mathbf{r}} \equiv 2\mathbf{r} \cdot \mathbf{v}.$$

On the right-hand side, $\mathfrak{R}$, of (105) we have

$$\mathfrak{R} = \sum_{i}{}' \sum_{j} \frac{C_{ij}(\mathbf{r}_i - \mathbf{r}_j) \cdot \mathbf{r}_i}{r_{ij}^3}. \tag{107}$$

In this expression i and j are dummy indices which we can replace by any other symbol without altering the value of $\mathfrak{R}$. Thus we have:

$$\mathfrak{R} = \sum_{i}{}' \sum_{j} \frac{C_{ji}(\mathbf{r}_j - \mathbf{r}_i) \cdot \mathbf{r}_j}{r_{ji}^3}. \tag{108}$$

But $r_{ij} = r_{ji}$ and, by (102), $C_{ij} = C_{ji}$, so we can rewrite (108) as

$$\mathfrak{R} = -\sum_i{}' \sum_j \frac{C_{ij}}{r_{ij}^3} (\mathbf{r}_i - \mathbf{r}_j) \cdot \mathbf{r}_j. \tag{109}$$

Adding one-half of (107) to one-half of (109) and using $(\mathbf{r}_i - \mathbf{r}_j) \cdot (\mathbf{r}_i - \mathbf{r}_j) = r_{ij}^2$, we obtain

$$\mathfrak{R} = \frac{1}{2} \sum_i{}' \sum_j \frac{C_{ij}}{r_{ij}} = U, \tag{110}$$

the potential energy, as in (103). The equality of $\mathfrak{L}$ and $\mathfrak{R}$ in (106*a*) and (110) gives:

$$-\frac{1}{2} \frac{d}{dt} (\Sigma\, M_i \mathbf{r}_i \cdot \mathbf{v}_i) + \begin{pmatrix}\text{total kinetic}\\ \text{energy}\end{pmatrix} = -\frac{1}{2} \begin{pmatrix}\text{total potential}\\ \text{energy}\end{pmatrix}. \tag{111}$$

The long-time average of the first term on the left-hand side of (111) is zero if the particles remain indefinitely within some finite volume, and we are left with the equality

$$\boxed{\langle \text{Total kinetic energy} \rangle = -\frac{1}{2} \langle \text{total potential energy} \rangle,} \tag{112}$$

for a collection of particles *bound* by electrostatic or gravitational forces.

When particles are confined by a container, the walls of the container exert additional forces which are not included in this derivation and which alter (112), sometimes in a rather subtle way. The result (112) remains valid even in situations where quantum corrections to the description of the behavior of particles cannot be ignored. We can exploit the virial theorem in discussion of electrons and nuclei in molecules or crystals, as well as of atoms bound to form a star or stars bound to form a galaxy.

The general virial theorem (112) implies not only that there must be kinetic energy when particles are held together by inverse-square-law forces but that the kinetic and potential energies are always comparable. Even if a group of particles is initially at rest, the attractive inverse-square forces will pull them together, increasing the magnitude of both the potential energy and the kinetic energy until ultimately the average kinetic energy comes to equal one-half the average potential energy. In the example below, we use the virial theorem to estimate the interior temperature of the sun, which like nearly all stars is a mass of dense hot gas.

The virial theorem is a clue to the internal structure of all matter in which inverse-square-law attractions contribute the dominant cohesive force. The average distance between atoms or nuclei in a *typical* star is probably always greater than 10^{-10} cm, corresponding to densities less than 10^7 gm/cm^3. The strong short-range nuclear forces have too small a range ($\sim 10^{-13}$ cm) to be effective at such distances; it is gravitational attraction that keeps the star together.

Example. Interior temperature of the sun. Estimate the average temperature of the interior of the sun. The gravitational self-energy U_s of a uniform star of mass M_s and radius R_s is

$$U_s = -\frac{3GM_s^2}{5R_s}, \tag{113}$$

as calculated earlier in this chapter. The average kinetic energy of a single atom in a star is proportional† to the absolute temperature T:

$$\langle \text{K.E. of a particle} \rangle = \tfrac{3}{2}kT, \tag{114}$$

with the constant k (the Boltzmann constant) given by:

$$k = 1.38 \times 10^{-16} \text{ erg/deg Kelvin.} \tag{115}$$

The total kinetic energy in the star is $\frac{3}{2}kNT_{\text{av}}$, where T_{av} is an appropriate average temperature over the interior of the star, and N is the number of atoms in the star. Then, using (113), the virial theorem gives:

$$\frac{3}{2}NkT_{\text{av}} = \langle \text{K.E. of all atoms} \rangle = -\frac{1}{2}\langle \text{P.E. of sun} \rangle \approx \frac{3GM_s^2}{10R_s}. \tag{116}$$

Thus we have

$$T_{\text{av}} \approx \frac{GM_s^2}{5R_sNk} = \frac{GM_sM}{5R_sk}, \tag{117}$$

where $M = M_s/N$ is the average mass of an atom in the star. Most of the atoms in a star are generally hydrogen or helium.

The mass of the sun M_s approximately equals 2×10^{33} gm, and the radius R_s is approximately 7×10^{10} cm. Let us take M as 3×10^{-24} gm, about twice the proton mass.‡ Then (117) becomes

$$T_{\text{av}} \approx \frac{(7 \times 10^{-8})(2 \times 10^{33})(3 \times 10^{-24})}{5(7 \times 10^{10})(1 \times 10^{-16})} \approx 10^7 \text{ deg Kelvin.} \tag{118}$$

This temperature corresponds to an energy of the order of 10^3 ev, which is enough to ionize completely atoms of low atomic number. Because the hydrogen and helium atoms are ionized, the number N of particles should be increased by the number of electrons, and, as we see from (117), the average temperature will be reduced to perhaps one-half or one-third of the value in (118). It is believed that the sun is not isothermal (at a uniform temperature) throughout its volume; our estimate is close to what sophisticated calculations suggest actually is the temperature of the core. The surface temperature is much lower; as estimated from the flux of radiation leaving the sun, the surface temperature is about 6×10^3 deg K. Our result (118) for the average temperature of the sun is over 10^3 times higher than the visually estimated surface temperature.

---

† This result is usually discussed in high school physics texts under the title *equipartition of energy;* we discuss this theorem in detail in Vol. V. The result is not always valid at very high densities because of quantum effects, but it is satisfactory under the condition prevailing in most hot stars.

‡ The abundance of hydrogen in the sun is believed to be about 60 percent by mass. Most of the rest is helium, with a little oxygen and traces of other elements.

This is a remarkable result which we have deduced from a little theory and only a small amount of experimental information, none of which involved removing ourselves from the earth. We cannot see into the sun's interior and yet we can calculate with some confidence the thermal conditions existing there. (The rate of energy radiation by the sun depends on the rate of burning of nuclear fuel in its core, and this rate furnishes an independent estimate of the core temperature.)

Example. Condensation of galaxies. Suppose we have a cloud of point particles with gravitational forces attracting each pair of particles and with an appreciable kinetic energy distributed among the particles. Under what conditions will the cloud expand, or contract, or maintain the same average size?

From the conservation of energy, the total kinetic and potential energy are related by

$$E = K + U. \tag{119}$$

The behavior of the cloud depends upon the initial relative magnitudes of the positive K.E. and the always negative P.E.:

(a) If K.E. $>$ |P.E.|, then the total energy $E > 0$ and the cloud (or at least part of it) will expand indefinitely. We prove that the cloud expands by showing that otherwise the virial theorem is contradicted. If the cloud remained within some finite volume, then the argument leading to the virial theorem is valid. According to this theorem $\langle \text{P.E.} \rangle = -2\langle \text{K.E.} \rangle$, so that

$$\langle E \rangle = \langle \text{K.E.} \rangle + \langle \text{P.E.} \rangle = -\langle \text{K.E.} \rangle \leqq 0, \tag{120}$$

where the angular brackets denote time average. But E is independent of time so $\langle E \rangle = E$ always; thus $E \leqq 0$ according to (120). The result contradicts the assumption that K.E. $>$ |P.E.|, and the cloud cannot remain within a finite volume; the cloud expands.

(b) If K.E. $<$ |P.E.|, then $E < 0$ and the bulk of the particles within the cloud cannot expand indefinitely. At infinite separations the potential energy becomes zero and all the energy is kinetic and hence positive. With an initially negative total energy such a state cannot be reached by all of the particles although, depending upon the initial conditions, some of them might escape to infinity. In general such a cloud of particles would be expected to hold together, because the virial theorem should be valid for the time-average quantities. We argue below that any sufficiently large cloud of gas for which $E < 0$ will ultimately condense to form a star. The argument does not constitute a firm proof that stars originate in this way but rather makes it plausible that they may have done so.

Effect of Radiative Energy Losses. Hot bodies lose energy in the form of electromagnetic radiation. As the temperature drops, there is a decrease in the rate of radiation of energy, but the rate never vanishes until the temperature reaches absolute zero. However, a cloud of gas held together by gravitational attraction cannot have zero temperature; actually *the temperature increases as the cloud radiates energy.*

This paradoxical-sounding result follows directly from the virial theorem and from the assumption of equipartition of energy. From (120) the total energy E is related to the time-average kinetic energy by

$$E = -\langle \text{K.E.} \rangle, \tag{121}$$

and from the statement (114) of equipartition of the energy we have

$$\langle \text{K.E.} \rangle = \tfrac{3}{2}NkT, \tag{122}$$

for N point masses in thermal equilibrium at temperature T. Thus

$$E = -\frac{3}{2}NkT; \qquad T = -\frac{2}{3Nk}E. \tag{123}$$

A variation ΔE in energy causes a variation ΔT in temperature, which, from (123), is

$$\Delta T = -\frac{2}{3Nk}\Delta E. \tag{124}$$

If ΔE is negative because of radiative losses, then ΔT must be positive. Further, by the virial theorem, the values of $\langle \text{K.E.} \rangle$ and $|\langle \text{P.E.} \rangle|$ must *increase* as E *decreases*. The increase in $|\langle \text{P.E.} \rangle|$ means that the gravitational binding increases and the cloud shrinks. In the same way a satellite which loses altitude because of friction will *gain* kinetic energy.

As energy is radiated away, the cloud gradually shrinks and *grows hotter;* as the cloud gets hotter, it both radiates and shrinks more rapidly, and its average temperature increases at a faster rate. Equation (117) relates the decreasing radius R_s of the star to its increasing average temperature T_{av}. Ultimately the temperature becomes so large that nuclear reactions can take place.† The gravitational collapse of the star slows down or stops when the nuclear reactions become the main source of energy, because the increased radiation pressure resists further contraction of the stellar matter.

---

† For nuclear reactions to occur, two nuclei have to get momentarily within about 10^{-13} cm of each other. Their Coulomb repulsion keeps the nuclei much further apart than this unless $T > 10^7$ deg K for protons and $T > 10^8$ deg K for helium nuclei. Quantitative estimates of these ignition or reaction temperatures can be made using quantum theory.

The spiral galaxy M81. (*Lick Observatory photograph*)

This is the present state of our sun. In about 7×10^9 years, when the hydrogen of the sun has largely been burned to helium, the collapse will begin again, and the average temperature of the sun's interior will resume its gradual increase.†

Any sufficiently large cloud of gas held together by its own gravitational attraction will become a star or stars in due course; this is a necessary consequence of an inverse-square-law attraction.

Now let us be speculative and try to apply our understanding of the nature of the gravitational attraction to the question of why the stars are clustered in galaxies. A typical galaxy may contain approximately 10^9 to 10^{11} stars. *Here we are on much less certain ground* than we were in our previous discussion. We shall attempt to extend a successful and undoubtedly correct model, the evolution of a star, to the origin of a galaxy, a problem much less well understood. This extension is bold and rash, in the style typical of initial attempts to interpret poorly understood phenomena in terms of known physical laws. (Much of the best pioneer work in the exact sciences is distinguished by daring approximations, simplifications, extrapolations.)

---

† In the event that, at some stage in the evolution of a star, (114) is not satisfied, then these qualitative considerations are no longer valid, because the kinetic energy will no longer be equal to $\frac{3}{2}kT$. As for normal solids and liquids, a group of particles bound by attractive forces may stop radiating and stop collapsing when quantum mechanical considerations dominate. In Vols. IV and V we shall learn to estimate the nature and importance of quantum effects in various situations.

From the fact that any mass of gas will contract if the magnitude of its potential energy appreciably exceeds that of its kinetic energy, we are led to the conclusion that any sufficiently large mass of gas with uniform density and temperature is gravitationally unstable. Different parts of the mass begin to collapse independently in large clumps if the magnitude of the potential energy of a clump exceeds twice the internal kinetic energy. A typical radius R for such a clump is given from (117) by

$$\frac{GM_c^2}{5R} \cong NkT, \tag{125}$$

where N is the number of particles in the clump, and M_c is the mass of the clump. In (125) the sign $>$ applies when the clump is in the process of collapsing and is not in the steady state envisaged by the virial theorem. If we assume the gas to be hydrogen atoms of mass M_{H}, then $M_c = NM_{\mathrm{H}}$. We also have $M_c = (4\pi/3)R^3 nM_{\mathrm{H}}$, where n is the concentration of hydrogen atoms. Thus (125) may be written as

$$R^2 > \frac{kT}{GnM_{\mathrm{H}}^2}, \tag{126}$$

where numerical factors of the order of unity have been dropped in the spirit of the inexactness of the estimate.

From (126) we obtain a remarkable result: A uniformly distributed gas of hydrogen atoms of density comparable to that presumed to exist in intergalactic space would spontaneously cluster into separate clouds comparable in mass to galaxies. Estimates of the concentration and temperature of intergalactic hydrogen give $n \simeq 10^{-5}$ cm^{-3} and $T \simeq 10^4$ deg K. In order that a part of a cloud be able to contract, the radius R must satisfy the inequality (126):

$$R^2 > \frac{(10^{-16})(10^4)}{(10^{-7})(10^{-5})(10^{-48})} \simeq \frac{10^{-12}}{10^{-60}} \simeq 10^{48}\ \mathrm{cm}^2, \tag{127}$$

or $R > 10^{24}$ cm. The mass of a cloud of this radius is

$$M_c \simeq (10^{24})^3(10^{-5})(10^{-24}) \simeq 10^{43}\ \mathrm{gm}. \tag{128}$$

From observation it is known that galaxies are separated by about 3×10^{24} cm, which is of the order of the initial radius estimated in (127). The radius of our galaxy is of the order of 10^{23} cm. The mass of the sun is 2×10^{33} gm, so the minimum mass we have calculated for a spontaneously contracting gas cloud is very roughly 5×10^9 solar masses. This number is of the right order of magnitude. The estimated masses of galaxies based upon observation, particularly of their luminosity, lie mostly in the range 10^9 to 10^{11} solar masses. This is in reasonable agreement with the conjecture that galaxies are spontaneous condensations caused by long-range gravitational attraction in a gas similar to that which may still exist in intergalactic space. Presumably conditions within the contracting galactic gas might later become appropriate for the formation of stars within the galaxy by a mechanism not too different from that which may have formed the galactic cloud itself.

The reader should note that we have *not* given a detailed, self-consistent dynamical theory of the origin of galaxies by condensation from a uniform gas. We have only made a suggestive estimate that leads to values of the galactic mass and intergalactic spacing which are not incompatible with inferences from astrophysical observations. The agreement in order of magnitude suggests either that the physical basis of the estimate is substantially correct or that we are victims of an unfortunate coincidence. We have not offered any explanation of the origin of the uniform distribution of gas. Further, the arguments we have given will not necessarily apply in an expanding universe because the kinetic energy of expansion may prevent condensation.

Advanced Topic 2: Numerical calculation of an orbit

The example is taken verbatim from the elementary physics lectures† of R. P. Feynman, with his kind permission and that of the copyright holders. These lectures by one of the great theoretical physicists of our time are penetrating, original, and amusing; the student will enjoy them.

". . . but can we analyze the motion of a planet around the sun? Let us see whether we can arrive at an approximation to an ellipse for the orbit. We shall suppose that the sun is infinitely heavy, in the sense that we shall not include its motion. Suppose a planet starts at a certain place and is moving with a certain velocity; it goes around the sun in some curve, and we shall try to analyze, by Newton's laws of motion and his law of gravitation, what the curve is. How? At a given moment it is at some position in space. If the radial distance from the sun to this position is called r, then we know that there is a force directed inward which, according to the law of gravity, is equal to a constant times the product of the sun's mass and the planet's mass divided by the square of the distance. To analyze this further we must find out what acceleration will be produced by this force. We shall need the *components* of the acceleration along two directions, which we call x and y. Thus if we specify the position of the planet at a given moment by giving x and y (we shall suppose that z is always zero because there is no force in the z-direction and, if there is no initial velocity

---

† R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman lectures on physics*, vol. I, sec. 9–7 (Addison-Wesley, Reading, Mass., 1963).

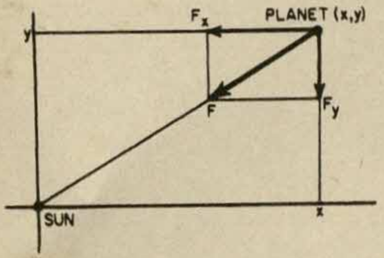


Fig. A. The force of gravity on a planet.

v_z, there will be nothing to make z other than zero), the force is directed along the line joining the planet to the sun, as shown in Fig. A.

"From this figure we see that the horizontal component of the force is related to the complete force in the same manner as the horizontal distance x is to the complete hypotenuse r, because the two triangles are similar. Also, if x is positive, F_x is negative. That is, $F_x/|F| = -x/r$, or $F_x = -|F|x/r = -GMmx/r^3$. Now we use the dynamical law to find that this force component is equal to the mass of the planet times the rate of change of its velocity in the x-direction. Thus we find the following laws:

$$m\frac{dv_x}{dt} = -\frac{GMmx}{r^3},$$

$$m\frac{dv_y}{dt} = -\frac{GMmy}{r^3}, \tag{129}$$

$$r = \sqrt{x^2 + y^2}.$$

This, then, is the set of equations we must solve. Again, in order to simplify the numerical work, we shall suppose that the unit of time, or the mass of the sun, has been so adjusted (or luck is with us) that $GM \equiv 1$. For our specific example we shall suppose that the initial position of the planet is at $x = 0.500$ and $y = 0.000$, and that the velocity is all in the y-direction at the start, and is of magnitude 1.6300. Now how do we make the calculation? We again make a table with columns for the time, the x-position, the x-velocity v_x, and the x-acceleration a_x; then, separated by a double line, three columns for position, velocity, and acceleration in the y-direction. In order to get the accelerations we are going to need Eq. (129); it tells us that the acceleration in the x-direction is $-x/r^3$, and the acceleration in the y-direction is $-y/r^3$, and that r is the square root of $x^2 + y^2$. Thus, given x and y, we must do a little calculating on the side, taking the square root of the sum of the squares to find r and then, to get ready to calculate the two accelerations, it is useful also to evaluate $1/r^3$. This work can be done rather easily by using a table of squares, cubes, and reciprocals: then we need only multiply x by $1/r^3$, which we do on a slide rule.

"Our calculation thus proceeds by the following steps, using time intervals $\epsilon = 0.100$: Initial values at $t = 0$:

$$x(0) = 0.500 \qquad y(0) = 0.000$$
$$v_x(0) = 0.000 \qquad v_y(0) = +1.630$$

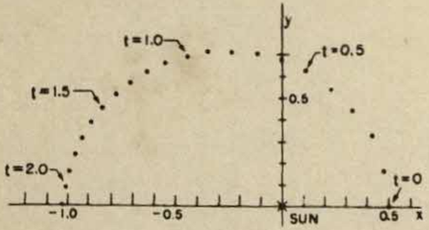


Fig. B. The calculated motion of a planet around the sun.

From these we find:

$$r(0) = 0.500 \qquad 1/r^3(0) = 8.000$$
$$a_x = -4.000 \qquad a_y = 0.000$$

Thus we may calculate the velocities $v_x(0.05)$ and $v_y(0.05)$:

$$v_x(0.05) = 0.000 - 4.000 \times 0.050 = -0.200;$$
$$v_y(0.05) = 1.630 + 0.000 \times 0.100 = 1.630.$$

Now our main calculations begin:

$$\begin{aligned}
x(0.1) &= 0.500 - 0.20 \times 0.1 = 0.480\\
y(0.1) &= 0.0 + 1.63 \times 0.1 = 0.163\\
r &= \sqrt{0.480^2 + 0.163^2} = 0.507\\
1/r^3 &= 7.67\\
a_x(0.1) &= 0.480 \times 7.67 = -3.68\\
a_y(0.1) &= -0.163 \times 7.70 = -1.256\\
v_x(0.15) &= -0.200 - 3.68 \times 0.1 = -0.568\\
v_y(0.15) &= 1.630 - 1.26 \times 0.1 = 1.505\\
x(0.2) &= 0.480 - 0.568 \times 0.1 = 0.423\\
y(0.2) &= 0.163 + 1.50 \times 0.1 = 0.313\\
&\qquad\text{etc.}
\end{aligned}$$

In this way we obtain the values given in [the table], and in 20 steps or so we have chased the planet halfway around the sun! In Fig. B are plotted the x- and y-coordinates given in [the table]. The dots represent the positions at the succession of times a tenth of a unit apart; we see that at the start the planet moves rapidly and at the end it moves slowly, and so the shape of the curve is determined. Thus we see that we *really do* know how to calculate the motion of planets!

"Now let us see how we can calculate the motion of Neptune, Jupiter, Uranus, or any other planet. If we have a great many planets, and let the sun move too, can we do the same thing? Of course we can. We calculate the force on a particular planet, let us say planet number i, which has a position x_i, y_i, z_i ($i = 1$ may represent the sun, $i = 2$ Mercury, $i = 3$ Venus, and so on). We must know the positions of all the planets. The force acting on one is due to all the other bodies which are located, let us say, at positions x_j, y_j, z_j. Therefore the equations are

$$\begin{aligned}
m_i \frac{dv_{ix}}{dt} &= -\sum_{j=1}^{N} \frac{Gm_im_j(x_i - x_j)}{r^3_{ij}},\\
m_i \frac{dv_{iy}}{dt} &= -\sum_{j=1}^{N} \frac{Gm_im_j(y_i - y_j)}{r^3_{ij}},\\
m_i \frac{dv_{iz}}{dt} &= -\sum_{j=1}^{N} \frac{Gm_im_j(z_i - z_j)}{r^3_{ij}}.
\end{aligned} \tag{130}$$

Further, we define r_{ij} as the distance between the two planets i and j; this is equal to

$$r_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}. \tag{131}$$

Also, Σ means a sum over all values of j —all other bodies—except, of course, for $j = i$. Thus all we have to do is to make more columns, *lots* more columns. We need nine columns for the motions of Jupiter, nine for the motions of Saturn, and so on. Then when we have all initial positions

Solution of $dv_x/dt = -x/r^3$, $dv_y/dt = -y/r^3$, $r = \sqrt{x^2 + y^2}$.
Interval: $\epsilon = 0.100$
Orbit $v_y = 1.63$ $v_x = 0$ $x = 0.5$ $y = 0$ at $t = 0$

t	x	v_x	a_x	y	v_y	a_y	r	$1/r^3$
0.0	0.500		−4.00	0.000		0.00	0.500	8.000
		−0.200			1.630			
0.1	0.480		−3.68	0.163		−1.25	0.507	7.675
		−0.568			1.505			
0.2	0.423		−2.91	0.313		−2.15	0.526	6.873
		−0.859			1.290			
0.3	0.337		−1.96	0.442		−2.57	0.556	5.824
		−1.055			1.033			
0.4	0.232		−1.11	0.545		−2.62	0.592	4.81
		−1.166			0.771			
0.5	0.115		−0.453	0.622		−2.45	0.633	3.942
		−1.211			0.526			
0.6	−0.006		+0.020	0.675		−2.20	0.675	3.252
		−1.209			0.306			
0.7	−0.127		+0.344	0.706		−1.91	0.717	2.712
		−1.175			0.115			
0.8	−0.245		+0.562	0.718		−1.64	0.758	2.296
		−1.119			−0.049			
0.9	−0.357		+0.705	0.713		−1.41	0.797	1.975
		−1.048			−0.190			
1.0	−0.462		+0.796	0.694		−1.20	0.834	1.723
		−0.968			−0.310			
1.1	−0.559		+0.858	0.663		−1.02	0.867	1.535
		−0.882			−0.412			
1.2	−0.647		+0.90	0.622		−0.86	0.897	1.385
		−0.792			−0.499			
1.3	−0.726		+0.92	0.572		−0.72	0.924	1.267
		−0.700			−0.570			
1.4	−0.796		+0.93	0.515		−0.60	0.948	1.173
		−0.607			−0.630			
1.5	−0.857		+0.94	0.452		−0.50	0.969	1.099
		−0.513			−0.680			
1.6	−0.908		+0.95	0.384		−0.40	0.986	1.043
		−0.418			−0.720			
1.7	−0.950		+0.95	0.312		−0.31	1.000	1.000
		−0.323			−0.751			
1.8	−0.982		+0.95	0.237		−0.23	1.010	0.970
		−0.228			−0.773			
1.9	−1.005		+0.95	0.160		−0.15	1.018	0.948
		−0.113			−0.778			
2.0	−1.018		+0.96	0.081		−0.08	1.021	0.939
		−0.037			−0.796			
2.1	−1.022		+0.95	0.001		0.00	1.022	0.936
		+0.058			−0.796			
2.2	−1.016		+0.96	−0.079		+0.07	1.019	0.945
					−0.789			
2.3								

Crossed x-axis at 2.101 sec, ∴ period = 4.20 sec.

$v_x = 0$ at 2.086 sec.

Cross x at 1.022, ∴ semimajor axis $= \dfrac{1.022 + 0.500}{2} = 0.761$.

$v_y = 0.796$.

Predicted time $\pi(0.761)^{3/2} = \pi(0.663) = 2.082$.

and velocities we can calculate all the accelerations from Eq. (130) by first calculating all the distances, using Eq. (131). How long will it take to do it? If you do it at home, it will take a very long time! But in modern times we have machines which do arithmetic very rapidly; a very good computing machine may take 1 microsecond, that is, a millionth of a second, to do an addition. To do a multiplication takes longer, say 10 microseconds. It may be that in one cycle of calculation, depending on the problem, we may have 30 multiplications, or something like that, so one cycle will take 300 microseconds. That means that we can do 3000 cycles of computation per second. In order to get an accuracy, of, say, one part in a billion, we would need 4×10^5 cycles to correspond to one revolution of a planet around the sun. That corresponds to a computation time of 130 seconds or about two minutes. Thus it takes only two minutes to follow Jupiter around the sun, with all the perturbations of all the planets correct to one part in a billion, by this method! (It turns out that the error varies about as the square of the interval ϵ. If we make the interval a thousand times smaller, it is a million times more accurate. So, let us make the interval 10,000 times smaller.)

"So, as we said, we began this chapter not knowing how to calculate even the motion of a mass on a spring. Now, armed with the tremendous power of Newton's laws, we can not only calculate such simple motions but also, given only a machine to handle the arithmetic, even the tremendously complex motions of the planets, to as high a degree of precision as we wish!"

Chapter 10

c as a fundamental constant of nature 311

Measurement of *c* 312

Transit time of light across orbit of the earth 312
Aberration of starlight 313
Toothed wheels and rotating mirrors 318
Cavity resonator 318
Shoran 319
Modulated light detector 320

Doppler effect 322

Source moving through medium; receiver at rest 323
Source at rest; receiver moving through medium 326
Source and receiver move together 326
Example: The recessional red shift 327

Speed of light in inertial frames in relative motion 328

Michelson-Morley experiments 332
Invariance of *c* 336
The ultimate speed 337

Film list 339

Further reading 339

Problems 339

Advanced Topic: Recoilless emission of gamma rays 342

The Speed of Light

c as a Fundamental Constant of Nature. The speed of light† in vacuum c is one of the fundamental constants of physics:

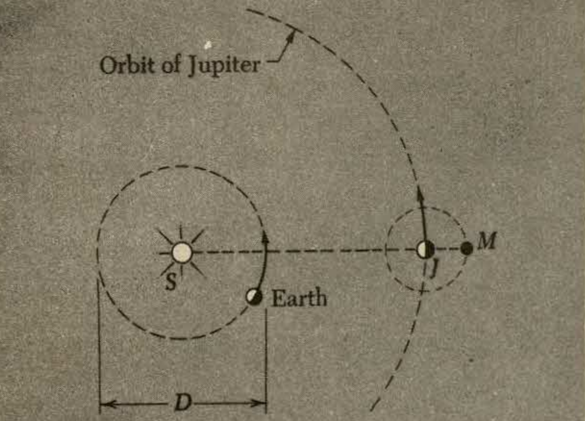
1. It is the speed at which all electromagnetic radiation travels in free space, independent of the frequency of the radiation.
2. No signal can be transmitted by any means whatsoever, in free space or in a material medium, at a speed faster than the speed of light c.
3. The speed of light in free space is independent of the reference frame from which it is observed. If the speed of a light signal is observed to be $c = 2.99793 \times 10^{10}$ cm/sec in one Galilean frame, it will be observed to be c and not $c + V$ (or $c - V$) in a second Galilean frame moving parallel to the signal with a speed V with respect to the first frame.
4. Maxwell's equations in electromagnetic theory and the Lorentz force equation involve the speed of light. This is apparent when they are written in Gaussian units.
5. The dimensionless constant (which is called the reciprocal of the fine-structure constant)

$$\frac{\hbar c}{e^2} \cong 137.04$$

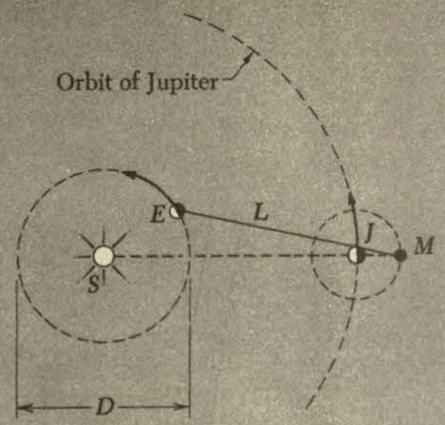
involves the speed of light. Here $2\pi\hbar$ is Planck's constant and e is the charge on the proton. This constant plays an important role in atomic physics and will be discussed in Vol. IV. We do not have a theory that predicts the value of this constant.

---

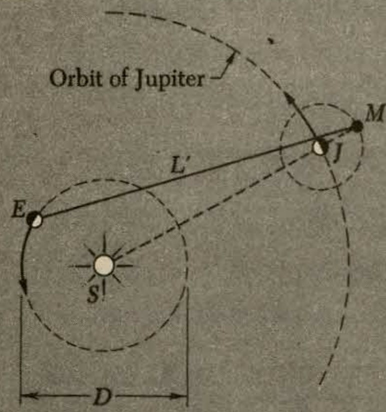
† Note that the phrase *speed of light* should always be understood to mean the speed of light in free space (c), unless it is explicitly stated otherwise. Thus the speed of light in a material medium is less than c and may even be less than the speed of a charged particle in the same medium (Čerenkov effect).



Eclipse of Jupiter's moon M occurs when Jupiter J is between sun S and M. This occurs about once every 42 hours, as M revolves around J.



Observation of eclipse at earth E occurs at time interval $\Delta t = L/c$ after actual eclipse, because of finite speed of light c.



An eclipse as observed at E about 6 months later. Now, $L' \approx L + D$, so $\Delta t' \cong L/c + D/c = \Delta t + D/c$. Roemer in 1667 measured $\Delta t' - \Delta t$, using the then current value of D to determine c.

This chapter is concerned chiefly with experiments and experimental results. We discuss the measurement of the speed of light and experimental evidence for the invariance of the speed of light with respect to the velocity of any inertial frame. We leave for Vol. III questions about the electromagnetic nature of light and the propagation of light in refractive and dispersive media such as solids and liquids. (A refractive medium is one in which the refractive index is not exactly unity. A dispersive medium is one in which the refractive index is a function of the frequency.)

Measurement of c

Many methods have been employed to determine the speed of light.† We list and sketch several of the methods here.

Transit Time of Light across Orbit of the Earth. For many centuries before there was experimental proof, it was believed that the speed of light must be finite. The first experimental evidence of the finite speed of light was due to Roemer in 1676. He observed that the motion of Io, the innermost moon of Jupiter, did not follow an entirely regular timetable. There was a variation in the periods of the eclipses of Io by Jupiter. The accumulated variation in the times of the eclipses of Io by Jupiter over one-half year of observation were of the order of 20 min. But this is about the transit time of light across the orbit of the earth: The average diameter D of the earth's orbit about the sun is 3×10^{13} cm

---

† An excellent review in English of measurements of tne speed of light is given by E. Bergstrand in *Handbuch der Physik*, S. Flügge (ed.), vol. 24, pp. 1–43 (Springer-Verlag OHG, Berlin, 1956). The values of c we quote are those listed by Bergstrand. See also J. F. Mulligan and D. F. McDonald, *Am. J. Phys.* **25**, 180 (1957).

(2 A.U.); the time required for light to cross from one side of the orbit to the other is then

$$T = \frac{D}{c} \cong \frac{3 \times 10^{13}}{3 \times 10^{10}} \cong 1000 \text{ sec} \cong 17 \text{ min}, \qquad (1)$$

which is in good agreement with the transit time of 16.6 min deduced from modern photometric observations of the same eclipses. In our estimate we have used the value 3×10^{10} cm/sec for the speed of light.

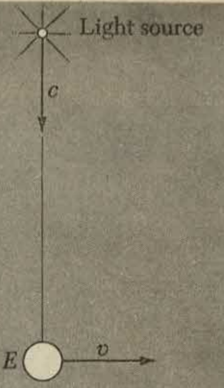
Roemer estimated 22 min for the value of the transit time deduced from the eclipses; he had available from other sources a rather inaccurate value of the diameter of the earth's orbit. He calculated from his data the value

$$c = 214{,}300 \text{ km/sec}. \qquad (2)$$

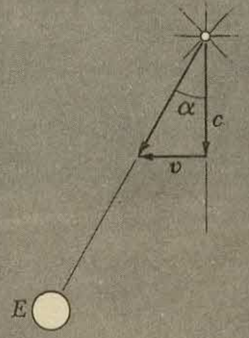
The angular motion of Jupiter about the sun is slower (12 years vs 1 year) than that of the earth; thus it is the diameter of the earth's orbit and not Jupiter's orbit which is chiefly involved in the calculation. The Roemer method is not very accurate, but it did show astronomers that in analyzing planetary observations to find the true motion of a planet or moon, it is necessary to make allowance for the propagation time of the light signal.

Aberration of Starlight. In 1725 James Bradley started an interesting series of precise observations of an apparent seasonal change in the position of stars, in particular of a star called γ Draconis. He observed that (after all other corrections had been applied) a star at the zenith (directly overhead) appeared to move in a nearly circular orbit with a period of a year, with an angular diameter of about 40.5″. He also observed that stars in other positions had a somewhat similar motion—in general, elliptical.

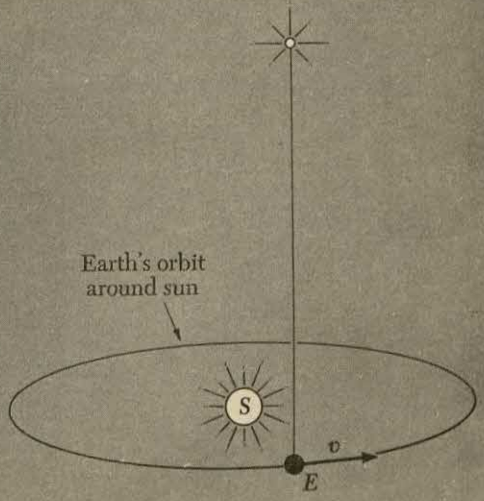
The phenomenon Bradley observed is called *aberration.* It has nothing to do with the true motion, if any, of the star; it arises from the finite speed of light and from the speed of the earth in its orbit about the sun. This was really the first direct experiment to suggest that the sun was a better inertial frame than the earth—i.e., that it is better to think of the earth as moving around the sun than of the sun as moving around the earth, for this experiment detects directly the annual change in the direction of the velocity of the earth relative to the stars.



Bradley in 1725 utilized the phenomenon of aberration to determine c. Suppose light from distant source illuminates object E, which has velocity v normal to incoming light.



According to an observer on E, light has horizontal velocity component v as well as vertical component c. Thus light ray from source is inclined at angle α, where $\tan \alpha = v/c$.



Bradley used light from distant star at zenith and known velocity of earth ($v_e = 30$ km/sec), to determine c from measurements of α: $\tan \alpha = v_e/c$.

A homely example of aberration: This student is caught in rain coming straight down. If he stands still under his umbrella, he keeps dry. But if he runs for it . . .

he gets wet. In his new reference frame the rain has horizontal velocity $-v$, where v is his velocity with respect to the ground.

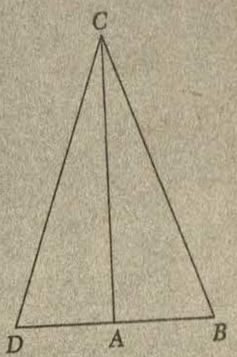
The simplest explanation of aberration is by the analogy of light propagation to the fall of raindrops. If no wind is blowing, raindrops fall vertically and a man at rest with an umbrella directly over his head does not get wet. If the man runs, holding the umbrella in the same position, the front of his coat will get wet. Relative to the moving person, the raindrops do not fall exactly vertically.

We quote from a pleasant account† of how the explanation of his observations came to Bradley: "At last, when he despaired of being able to account for the phenomena which he had observed, a satisfactory explanation of it occurred to him all at once, when he was not in search of it.‡ He accompanied a pleasure party in a sail upon the river Thames. The boat in which they were was provided with a mast, which had a [weather] vane at the top of it. It blew a moderate wind, and the party sailed up and down the river for a considerable time. Dr. Bradley remarked that, every time the boat put about, the vane at the top of the boat's mast shifted a little, as if there had been a slight change in the direction of the wind. He observed this three or four times without speaking; at last he mentioned it to the sailors, and expressed his surprise that the wind should shift so regularly every time they put about. The sailors told him that the wind had not shifted, but that the apparent change was owing to the change in the direction of the boat, and assured him that the same thing invariably happened in all cases. This accidental observation led him to conclude, that the phenomenon which had puzzled him so much was owing to the combined motion of light and of the earth."

---

† T. Thomson, *History of the Royal Society*, p. 346 (London, 1812).

‡ Many inventions and discoveries are made when, after an initial failure, the scientist has taken his thoughts away from the problem. A distinguished mathematician discusses this effect in a fascinating and important little book: J. Hadamard, *An essay on the psychology of invention in the mathematical field* (Princeton Univ. Press, Princeton, N.J., 1945); reprint (Dover, New York, 1954).



This is the explanation of aberration in Bradley's words:† "I considered this matter in the following manner. I imagined *CA* [see the figure] to be a ray of light, falling perpendicularly upon the line *BD*; then if the eye is at rest at *A*, the object must appear in the direction *AC*, whether light be propagated in time or in an instant. But if the eye is moving from *B* towards *A*, and light is propagated in time, with a velocity that is to the velocity of the eye, as *CA* to *BA*; then light moving from *C* to *A*, whilst the eye moves from *B* to *A*, that particle of it by which the object will be discerned when the eye is in motion comes to *A*, is at *C* when the eye is at *B*. Joining the points *B*, *C*, I supposed the line *CB* to be a tube (inclined to the line *BD* in the angle *DBC*) of such a diameter as to admit of but one particle of light; then it was easy to conceive that the particle of light at *C* (by which the object must be seen when the eye, as it moves along, arrives at *A*) would pass through the tube *BC*, if it is inclined to *BD* in the angle *DBC*, and accompanies the eye in its motion from *B* to *A*; and that it could not come to the eye, placed behind such a tube, if it had any other inclination to the line *BD*."

For a star directly overhead the maximum aberration occurs when the earth's velocity is perpendicular to the line of observation. Then the tilt angle, or aberration, of the telescope is seen from the figures to be given by

$$\tan \alpha = \frac{v_e}{c}, \tag{3}$$

where v_e is the speed of the earth. The orbital speed of the earth about the sun is 3.0×10^6 cm/sec; the speed due to rotation about the earth's own axis, which is about 100 times slower, may be neglected here. Thus

$$\tan \alpha \cong \frac{3.0 \times 10^6}{3.0 \times 10^{10}} \cong 1.0 \times 10^{-4} \text{ radians} \cong \alpha, \tag{4}$$

---

† J. Bradley, *Phil. Trans. Roy. Soc.* London 35, 637 (1728).

because $\alpha \cong \tan \alpha$ for small α. Measured in seconds of arc,

$$\alpha = 20.5''. \tag{5}$$

Twice this, or 41″, is to be compared with the value of 40.5″ observed by Bradley for the angular diameter of the apparent stellar orbit.

Bradley's telescope is shown in the figure. It was about twelve feet long and was designed specifically for the accurate observation of stars near the zenith. The agreement of his observations on γ Draconis with his hypothesis is shown in the following table taken from the original paper.

In Chap. 11 we shall reconsider the Bradley experiment from the viewpoint of special relativity, and we shall confirm (3) to an accuracy of the order of α. (Often we find that effects calculated to first order in v/c are not affected by relativity. This is not always so, however.)

			The difference of declination by observation, sec of arc	*The difference of declination by the hypothesis, sec of arc*
1727.	October	20	4½	4½
	November	17	11½	12
	December	6	17½	18½
	December	28	25	26
1728.	January	24	34	34
	February	10	38	37
	March	7	39	39
	March	24	37	38
	April	6	36	36½
	May	6	28½	29½
	June	5	18½	20
	June	15	17½	17
	July	3	11½	11½
	August	2	4	4
	September	6	0	0

Plate A....at the end of the memoirs.

⊙ The end of the axis with dot for the plumb line.
a. The Y on which it rests.
c. The screw for regulating the plumb line.
b.b. Adjusting screws.
d.d. Iron supporters.
i.k. Braſs supporter.
e. Screw & support for fastening the wood guard to the Braſs supporter i.k.
f. The Micrometer screw.
g. A screw for taking off the preſsure of the telescope from the micrometer screw when the instrument is not in use.
NB The telescope is put out of the vertical position in order to shew the wood guard for the plumb line. And the arc to the right of the telescope is broken off in order to shew the micrometer screw.
h. A back support on which the micrometer slides and to which it is clamped.

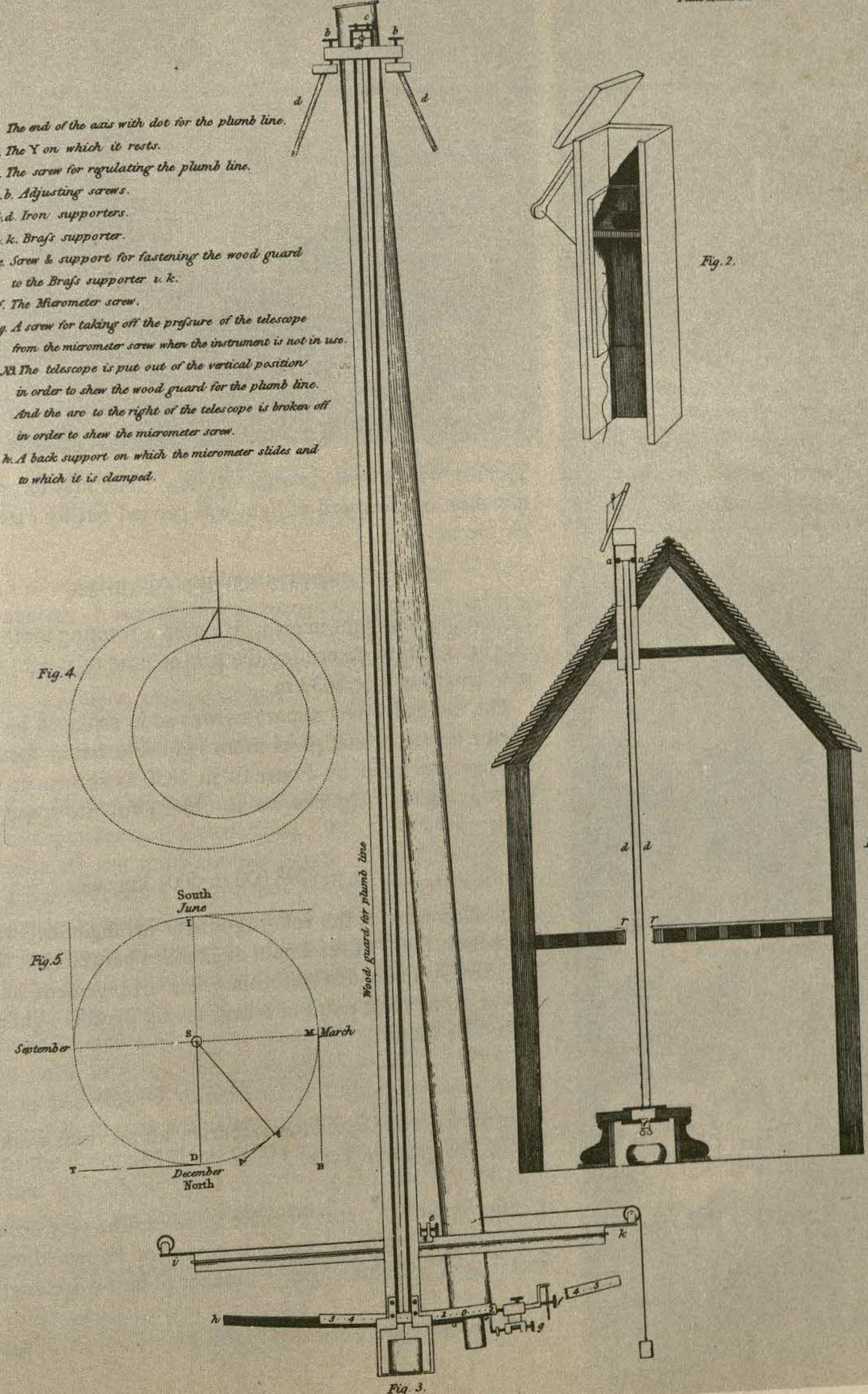


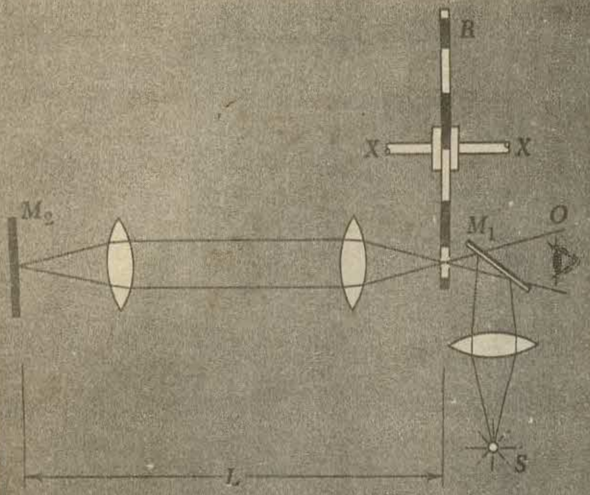
Fig. 3.

Fig. 2.

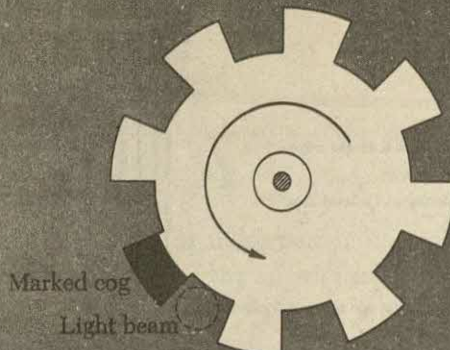
Fig. 1.

Fig. 4.

Fig. 5.



Fizeau's cogwheel apparatus, 1849. Light from point source S is reflected from half-silvered mirror M_1 past cogwheel R rotating on axis X-X. Light then goes to mirror M_2 and returns to observer O through R and M_1.



View of light beam and cogwheel R seen by observer O. Rotation of R chops light beam from S, M_1 into short pulses. (Light can only pass from M_1 to M_2 if no cog is in the way.)

Toothed Wheels and Rotating Mirrors. The first terrestrial determination of the speed of light was carried out by Fizeau in 1849. He found

$$c = (315{,}300 \pm 500)\ \text{km/sec}$$

for the speed of light in air.† He used a rotating toothed wheel as a light switch to determine the transit time of a light flash over a path length of 2 × 8633 m.

The toothed-wheel apparatus was soon replaced by a rotating-mirror device, which gives more light and better focusing. The arrangement used by Foucault in 1850 is shown in the accompanying figure. His best value (1862) for the speed of light in air is

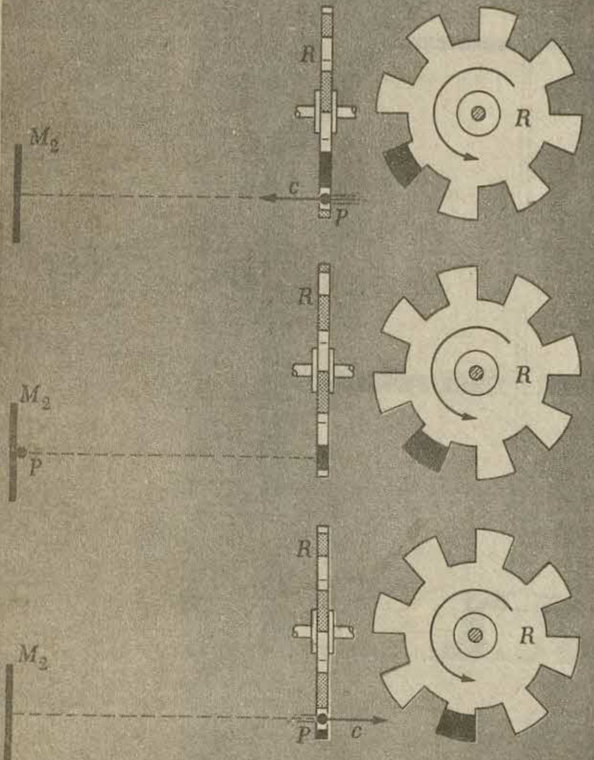
$$c = (298{,}000 \pm 500)\ \text{km/sec}.$$

A development of the rotating-mirror arrangement was used by Michelson (1927) over a path of 22 miles between Mt. Wilson and Mt. San Antonio in California. His arrangement has the light source at the focal point of a lens, giving parallel light over a long path. He found

$$c = (299{,}796 \pm 4)\ \text{km/sec}.$$

This work greatly exceeds in accuracy all previous work. Further details are given in Prob. 3.

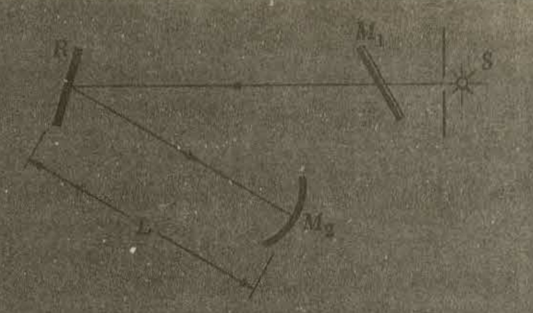
Cavity Resonator. It is possible to determine very accurately the frequency at which a resonant cavity of known dimensions (a metal box) contains a known number of half-wavelengths of elec-



Pulse P with velocity c must travel to M_2 and return to R (total distance $2L$) in time during which cogs move over one space, if pulse is to be transmitted to O. Fizeau determined c from L and angular velocity of R.

---

† The speed of light in vacuum is calculated to be about 91 km/sec faster than in air.



Foucault's rotating mirror apparatus, 1850, consisting of source slit S, half-silvered mirror M_1, rotating mirror R (axis of rotation normal to page), and spherical mirror M_2. Beam path from S to M_2 shown.

tromagnetic radiation. The speed of light is then calculated from the theoretical relation

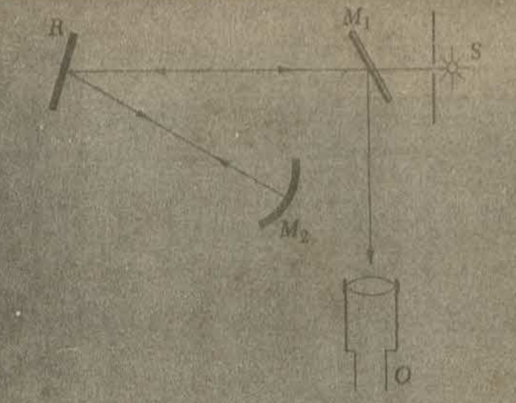
$$c = \lambda\nu \tag{6}$$

connecting the wavelength λ and the frequency ν. The cavity is usually evacuated. It is necessary to correct the inside dimensions of the cavity for the small penetration† of the electromagnetic field into the surface of the metal. Essen (1950) used frequencies of 5960, 9000, and 9500 Mc/sec to find

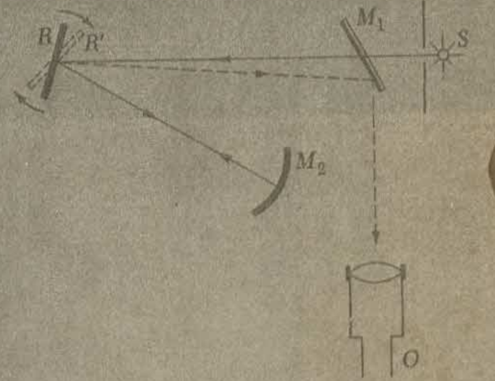
$$c = (299{,}792.5 \pm 1)\ \text{km/sec}.$$

Shoran.‡ The Shoran method utilizes a radar beacon. A radar beacon or transponder is a device which, when it receives a radar pulse, immediately transmits another pulse. In effect it is somewhat like a nondirectional mirror, but a mirror which amplifies the incoming signal before reemitting it.

In the measurement of the speed of light, radar beacons are put at points A and B. A radar transmitter located somewhere on the line between the two points emits pulses of electromagnetic energy and a receiver located at the same place measures the time for a transmitted pulse to reach one of the beacons and be returned. The distances between the transmitter and radar beacons can be measured very accurately with standard surveying methods. For distances of the order of 10^7 cm (100 km), the error is only about 10 cm. Time intervals as short as 10^{-9} sec can also be measured very accurately with quartz-crystal clocks.



With R stationary, light beam from M_1 to R to M_2 is reflected back along same path to M_1, detected by O.

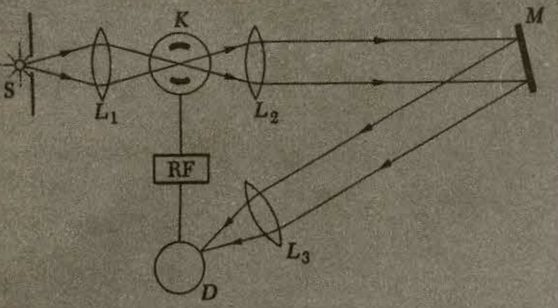


If mirror R rotates, light from S to R to M_2 returns when rotating mirror is in new position R'. Thus O observes displaced image on M_1. Foucault determined c from L, image displacement, and mirror angular velocity.

---

† The penetration region is known as the *skin depth*. It is of the order of 1 micron (abbreviated μ; 1 $\mu \equiv 10^{-4}$ cm) in thickness in copper at room temperature at 10^{10} cps. There are also other corrections to be applied.

‡ This word denotes SHOrt RAnge Navigation.



A modern method for determination of c. Light from source S is amplitude-modulated in Kerr cell K, then passed to mirror M and the photoelectric detector D through lenses $L_{2,3}$. The photodetector sensitivity and Kerr cell are synchronized by modulated radio-frequency voltage generator RF.

If t_A is the time required for a pulse to travel from the transmitter to the beacon at A, then the measured elapsed time will be

$$T_A = 2t_A + \delta_A, \tag{7}$$

where δ_A is the time delay in the response of the beacon. The time delay is easily determined in a separate measurement. We could measure the velocity of light by locating the transmitter at point B and calculating c from the relation

$$c = \frac{2L_{AB}}{2t_A}, \tag{8}$$

where L_{AB} is the distance between A and B. But we can accumulate much more data and make use of a longer line-of-sight if we locate the pulse transmitter on an aircraft flying at high altitude and measure the delay of signals returned from the transponders at both A and B. The aircraft flies over the line between them. In this way Aslakson obtained

$$c = (299{,}794.2 \pm 1.9) \text{ km/sec.}$$

The accuracy of the method is limited by the measurement of ground distances and by atmospheric conditions, but the value of c obtained is very good. The original work exposed systematic errors in earlier measurements.

Modulated Light Detector. A mirror M reflects light from a source S to a photocell detector D. The intensity of the source is modulated by an oscillator at a radio frequency; the oscillator also modulates the sensitivity of the photocell at the same frequency. The response will be a maximum if light of maximum intensity reaches the photocell at a time of maximum sensitivity. For this to happen, the time for light to travel from S to D must be equal to an integral number of periods N of the modulating radio frequency ν. Then the elapsed time is N/ν, from which

$$c = \frac{2L\nu}{N}, \tag{9}$$

if L is the distance from S and D to the mirror. The path lengths employed are of the order of 10 km.

Using this method Bergstrand measured

$$c = (299{,}793.1 \pm 0.3) \text{ km/sec.}$$

Note that the estimated error is very low. The same device is

The intensity of light entering the Kerr-cell system from the source is steady, . . .

but the light coming out of the Kerr-cell system is modulated. The time of transit of the light from K to D can be varied by moving M: M can be adjusted so light arrives at D as shown.

If we move M out a little, the light arrives later . . .

for M *further* out, light arrival is *still* later . . .

for M *further* out, light arrival is *still* later.

for M further out, light arrival is still later . . .

Now suppose the sensitivity of the detector is modulated as shown here . . .

The detector responds only when it is sensitive and when light is coming in.

Thus we have this detector response for condition a.

For condition b we have this: The incoming light and the detector sensitivity are in phase.

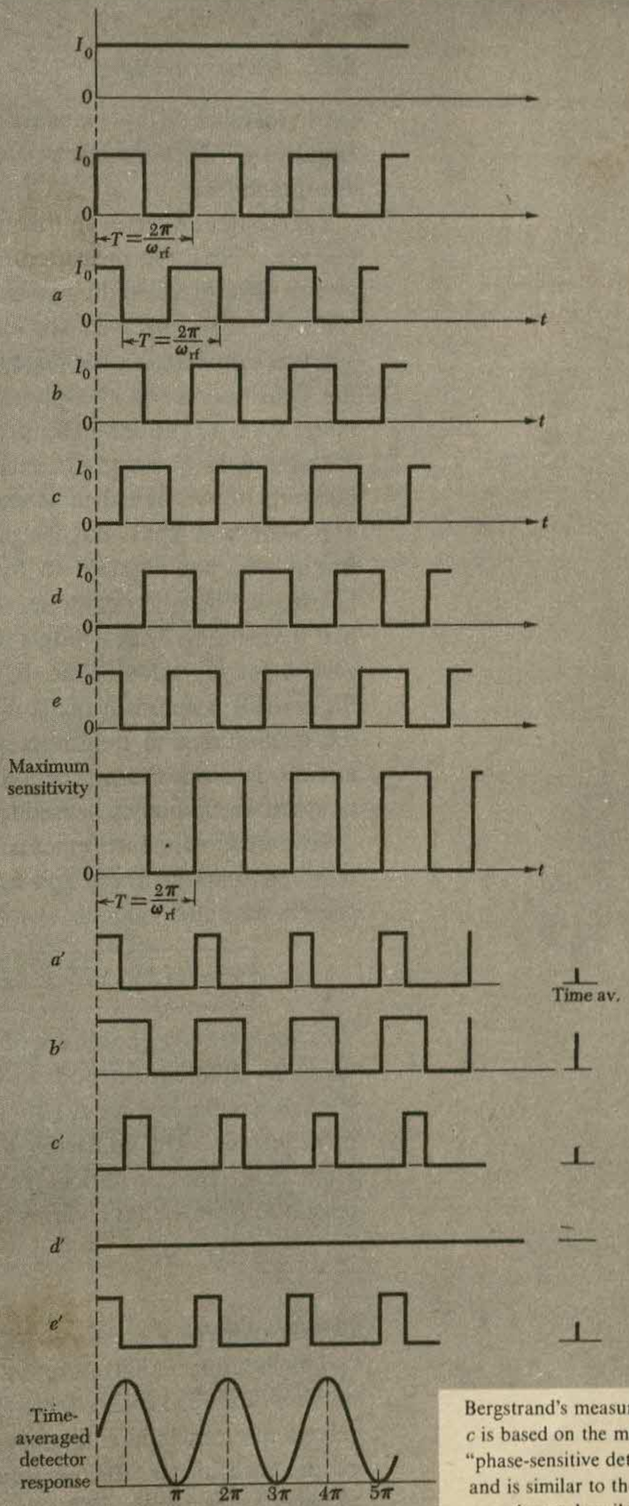
For c we have this.

For d the arriving light and the detector sensitivity are 180° out of phase, so there is no response.

For e we have this.

As we continuously vary the position of M, we obtain this average detector response.

The distance between two successive maxima of this curve corresponds to a change in the light path $2\Delta L$ caused by displacement of M:


I_0
0
$T = \frac{2\pi}{\omega_{rf}}$
a
b
c
d
e
t
Maximum sensitivity
a'
b'
c'
d'
e'
Time av.
Time-averaged detector response
π
2π
3π
4π
5π

Bergstrand's measurement of c is based on the method of "phase-sensitive detection" and is similar to the experiment described here.

$$\frac{2\Delta L}{c} = \frac{1}{\nu_{rf}};\ \therefore c = 2\Delta L \nu_{rf}$$

used (together with a standard value for c) to determine geodetic lengths over distances up to 40 km; in this application it is known as a *geodimeter.*

Two sources of error in this type of experiment are of particular interest. First, the measured time interval includes not only the transit time of the light but also the transit time for electrons to carry the signal between the electrodes in the photocell detector. The electron transit time depends on the location of the image of the light source on the photocathode. A displacement of the image by a few millimeters produces a difference in transit time of as much as 10^{-9} sec. In early experiments of this type the time intervals of two light beams were compared. The path length of one beam was fixed, and the path length of the other was varied. But it was not possible to focus coincident images from both beams on the photocathode. Using a single beam, Bergstrand had to deal only with a single image, and thus although he had to correct for the transit time, by proper focusing he could arrange for it to be a constant of the apparatus. Second, the conditions for maximum and minimum photocell response are rather uncritical, because the top of a sine wave is not sharp. A second modulation frequency is used to define the response more sharply.

Hundreds of measurements of c have been made in the past hundred years by these and a dozen or so other methods. The present accepted value is

$$\boxed{c = (2.997\,925 \pm 0.000\,003) \times 10^{10}\ \text{cm/sec.}} \qquad (10)$$

This represents a consensus of the most reliable recent measurements by different methods in which electromagnetic waves from 10^8 cps (radio frequency) to 10^{22} cps (gamma rays) have been investigated. The precision at the highest frequency is not as great as at radio or optical frequencies, but there is at present no reason to believe that c varies with the frequency of the radiation.

Doppler Effect

This is a convenient point at which to give a nonrelativistic discussion of the Doppler effect. The Doppler effect or Doppler shift relates the measured frequency of a wave to the relative velocities of the transmitter, the medium, and the receiver. It is advantageous to begin the discussion with sound waves propagating in a liquid or gas, because the role of the medium is intuitively clear for sound waves. We know that there cannot be sound waves without a medium in which they can propagate. For light waves we shall see that there is no medium: this is the conclusion from experiment.

Source Moving through Medium; Receiver at Rest. Let the transmitter be fixed at the origin of a Galilean reference frame T which moves with respect to the receiver located at the origin of another Galilean reference frame R. We suppose for the present that the propagation medium M is at rest with respect to R, so that the frame R is identical with the frame M. The transmitter frame T moves with respect to R (and M) with a velocity

$$\mathbf{V} = V\hat{\mathbf{x}}. \tag{11}$$

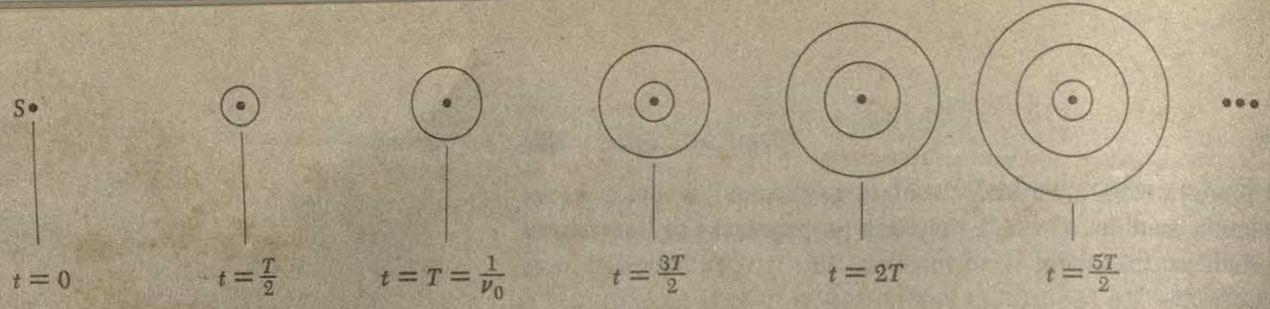
For V positive the transmitter moves from the left toward the receiver.

The velocity of the sound *relative* to the medium depends only on the mechanical properties of that medium and not at all on the velocity of the source relative to the medium. It is something like throwing objects on a conveyor belt. No matter how fast you run parallel to the belt, the velocity of the object once it lands on the belt (and sticks) will be just the velocity of the conveyor belt itself. Given a specific medium, the velocity of sound in that medium is determined. We will have the following relationship between the wavelength, frequency, and velocity of the wave:

$$\lambda_R \nu_R = v_s. \tag{12}$$

This relationship is obvious provided v_s is a constant, as ν_R is the number of cycles that pass a point per second and λ_R is simply the length (in centimeters, say) of each of these cycles. Thus the number of centimeters that pass the point per second is given by $\lambda_R \nu_R$. The density and elastic constants of the medium determine the velocity of the sound wave. If we generate a frequency ν_R, the wavelength λ_R is uniquely determined by (12). (Elastic waves are discussed in Vol. III.)

Suppose that the transmitter emits in the $+\hat{\mathbf{x}}$ direction a train of N waves in time t. (By one wave we mean one complete wavelength.) The first wave will advance in the medium a distance $v_s t$ in this time; the last wave will have just left the transmitter at the end of this time, and the transmitter will have advanced a distance Vt with respect to the medium. The distance between the front and the rear of the wave train is $(v_s - V)t$, and there are N



Sound waves generated by stationary source:
S emits spherical sound waves, wave length λ_0,
frequency ν_0, velocity $v_s = \nu_0\lambda_0$.

waves in this distance. Thus the wavelength is

$$\lambda_R = \frac{(v_s - V)t}{N}, \tag{13}$$

and the frequency as seen in the medium or as received in R is

$$\nu_R = \frac{v_s}{\lambda_R} = \frac{N}{t}\,\frac{v_s}{v_s - V}. \tag{14}$$

The transmitter, however, emits N waves in t sec, so its frequency in its own reference frame is

$$\nu_T = \frac{N}{t}. \tag{15}$$

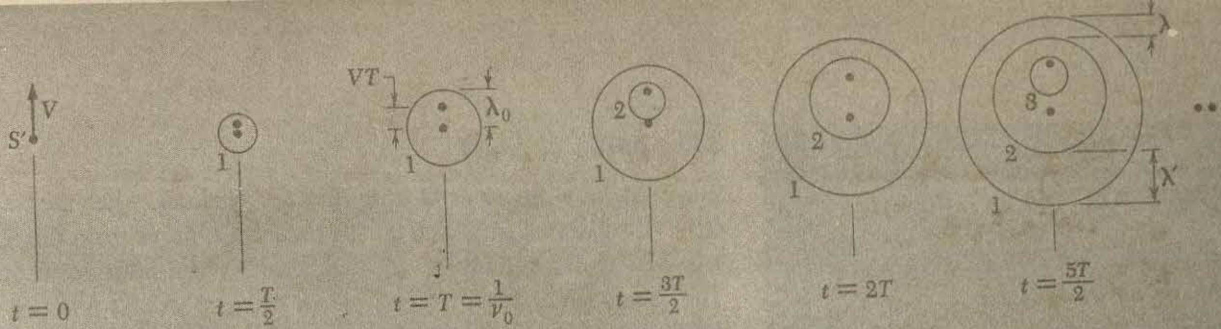
On substitution of (15) in (14) we have

$$\boxed{\nu_R = \frac{\nu_T}{1 - V/v_s}.} \tag{16}$$

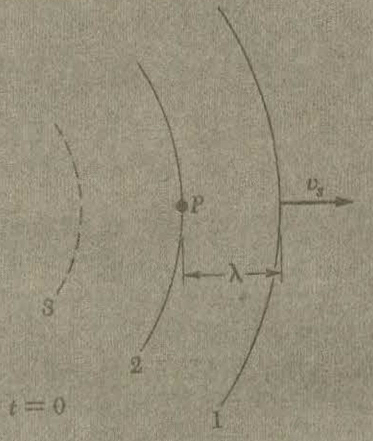
Thus if V is positive (transmitter moving toward the receiver), the frequency perceived by the receiver is higher than the frequency transmitted. If V is negative (transmitter moving away from receiver), the frequency received is lower than the frequency transmitted. The shift in frequency is known as the *Doppler effect* or *Doppler shift.* For a jet aircraft, V is of the same order of magnitude as the velocity of sound in air, and the Doppler effect is quite large. If $V/v_s \ll 1$ we may approximate (16) by

$$\nu_R \cong \nu_T\left(1 + \frac{V}{v_s}\right), \tag{17}$$

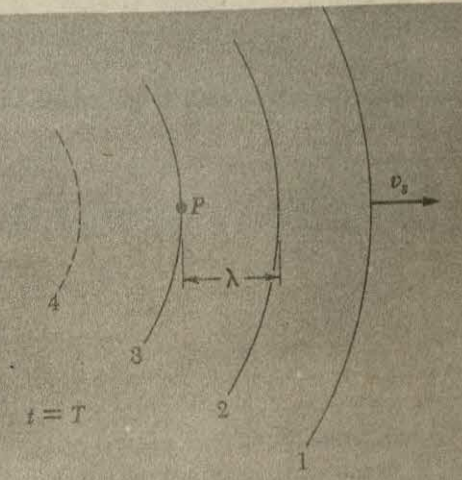
to terms of order V/v_s.



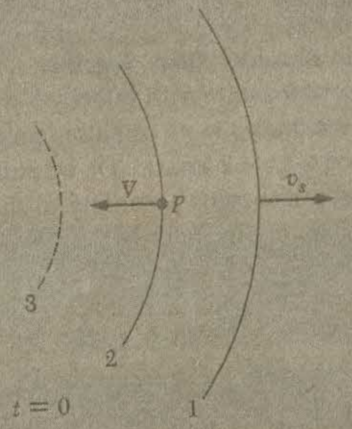
Doppler effect: Source S′ has velocity $\dot{y} = V$ with respect to medium and emits waves of frequency ν_0. Wavefront 1 is emitted when S′ is at $y = 0$ and propagates radially outward from there. Wavefront 2 is emitted when S′ is at $y = VT$ and propagates radially outward from there, etc.



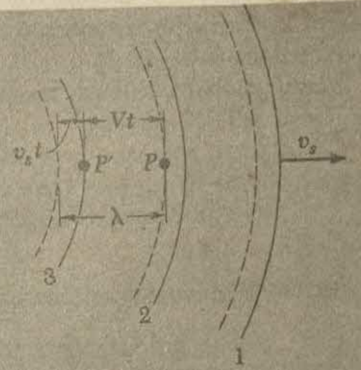
An *observer* stationary in the medium at point *P* observes waves of wavelength λ.



The frequency of these waves, according to observer at *P*, is $\nu = v_s/\lambda$.

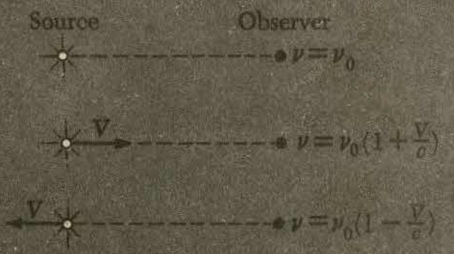


Now consider an observer moving in medium with velocity V. Suppose wavefront 2 is at *P* at $t = 0$.



Then observer will reach next wavefront when $(v_s + V)t = \lambda$. Thus to moving observer,

$$\nu_R = \nu\left(1 + \frac{V}{v_s}\right).$$


Source
Observer
$\nu = \nu_0$
V
$\nu = \nu_0(1 + \frac{V}{c})$
V
$\nu = \nu_0(1 - \frac{V}{c})$

For light waves this relation still holds in the approximation $V \ll c$, but we no longer may ask whether it is the source or observer which moves.

Source at Rest; Receiver Moving through Medium. Let the transmitter T be at rest in M and let the receiver R move from the right toward T with a velocity $\mathbf{V} = -V\hat{\mathbf{x}}$, where V is a positive number. The wavetrain of N waves emitted in time t occupies a length v_st in the medium. The velocity of the sound wave relative to the receiver is $v_s + V$, so that the receiver sees N waves in a time

$$\frac{v_st}{v_s + V}.$$

The frequency seen by R is therefore

$$\nu_R = \frac{N}{v_st/(v_s + V)} = \frac{N}{t} \cdot \frac{v_s + V}{v_s}. \tag{18}$$

Now $\nu_T = N/t$, so that

$$\nu_R = \nu_T\left(1 + \frac{V}{v_s}\right). \tag{19}$$

This is not exactly the same as the result (16) for the earlier problem, but we see from a comparison of (17) with (19) that the two results are identical to terms of first order in V/v_s. We shall see in Chap. 11 that the results (16) and (19) are valid also for light waves in free space, but *only* to first order in V/c. For sound waves the terms of second order in V/c are different in (17) and (19), so for sound waves we are able to tell by experiment whether the transmitter or the receiver is moving with respect to the medium. The medium is essential for the propagation of sound waves. For light waves it will turn out that the second-order terms are equal.

Source and Receiver Move Together. Suppose both source and receiver move together with velocity V with respect to the medium. We easily see that $\nu_R = \nu_T$, as pulses emitted a time $T = 1/\nu_T$ apart are received a time T apart. Or we can combine (16) and (19), using appropriate signs according to the direction of the motion of the transmitter and the receiver:

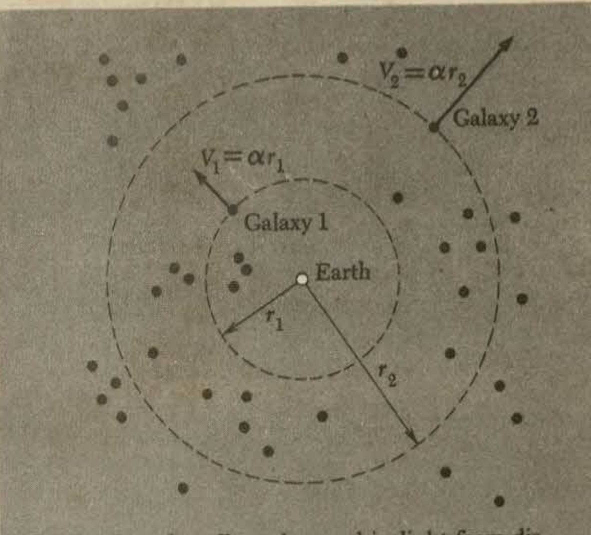
$$\nu_R = \frac{1}{1 \pm V/v_s} \cdot \left(1 \pm \frac{V}{v_s}\right)\nu_T = \nu_T. \tag{19a}$$

Thus the frequency is unchanged; the apparent velocity of propagation is changed, of course, by the flow of the medium.

Example. The recessional red shift. Spectrographic analysis of light received from distant galaxies shows that certain prominent spectral lines are shifted very significantly toward the red end or low-frequency end of the visible spectrum. This shift may be interpreted as a Doppler shift arising from the velocity of recession of the source. It is also known that the fractional magnitude $\Delta\nu/\nu$ or $\Delta\lambda/\lambda$ of the shift is directly proportional to the distance of the source from us, provided that the fractional shift is $\ll 1$. This is an extraordinary and provocative observational fact. The simplest nonrelativistic explanation of the distance-velocity relation is known as the "big-bang" theory, according to which the universe was formed from an explosion about 10^{10} years ago. The fastest-moving products of the original explosion now form the outermost regions of the universe. Thus the greater the radial velocity of matter (relative to us), the farther it is from us and the greater is its red shift. There also are more sophisticated explanations of the recessional red shift. None is proved.

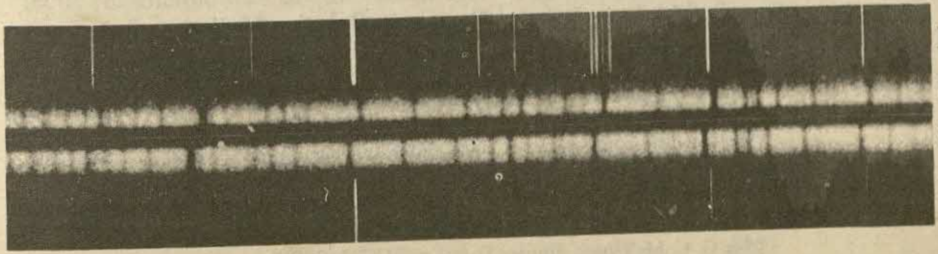
A pair of easily recognizable absorption lines in the spectrum of potassium (the K and H lines) are prominent in the spectra of many stars. These lines occur near wavelength† 3950 Å in laboratories on earth. We assume that laboratory observers moving in the rest frame of any star would measure the same wavelength. In light coming from a nebula in the constellation Boötes we observe these same lines at a wavelength of 4470 Å, a shift toward the red of 4470 − 3950 = 520 Å. This is a relative shift of

$$\frac{\Delta\lambda}{\lambda} = \frac{520}{3950} = 0.13. \tag{20}$$



The Doppler effect observed in light from distant stars indicates that the galaxies are receding from us with a velocity proportional to their distance from Earth.

Two spectrograms (taken at different times) of the binary star α^1 Geminorum. Only one of the two stars in this binary emits enough light to be detected. Notice that the spectral lines from the star are shifted, with respect to the laboratory reference lines, in different directions corresponding to two phases of motion of the star. In one phase the star is moving toward the earth and the frequency of the light is increased; in the other phase the star is moving away from the earth and the frequency is decreased. (*Lick Observatory photograph*)



---

† 1 angstrom ≡ 10^{-8} cm.

We observe that, by using (16) with v_s equal to c (as will be justified for light waves in Chap. 11), and differentiating† $\nu = c/\lambda$ with c constant,

$$\frac{\Delta\nu}{\nu} = -\frac{\Delta\lambda}{\lambda}, \qquad \text{or} \qquad \frac{\Delta\lambda}{\lambda} \cong -\frac{V}{c}, \tag{21}$$

where we have used (17) with v_s equal to c, as will be justified for light waves in Chap. 11. We infer from (21) that the nebula is receding from us with a relative speed $|V| \cong 0.13c$, which is really quite fast. For *higher speeds* we need to use one or another relation for the Doppler shift as modified by the theory of relativistic models of the universe.‡

Similar observations on large numbers of galaxies can be combined with independent estimates of their distances to obtain an amazing empirical result: The relative velocity of a galaxy at distance r from us may be represented by the relation

$$V = \alpha r, \tag{22}$$

where the constant α is empirically determined to be about 3×10^{-18} sec^{-1}. (The estimation of galactic distances is a complex subject for which an astronomy text must be consulted.) The reciprocal of α has the dimensions of time:

$$\frac{1}{\alpha} \cong 3 \times 10^{17} \text{ sec} \cong 10^{10} \text{ years}. \tag{23}$$

When we multiply $1/\alpha$ by c, we obtain a length:

$$\frac{c}{\alpha} \cong (3 \times 10^{10})(3 \times 10^{17}) \cong 10^{28} \text{ cm}. \tag{24}$$

The time (23) is loosely called the "age of the universe"; the length (24) is loosely called the "radius of the universe." The real significance of these quantities is not known at present, although several different cosmological models have been proposed to account for the form of the relation (22).

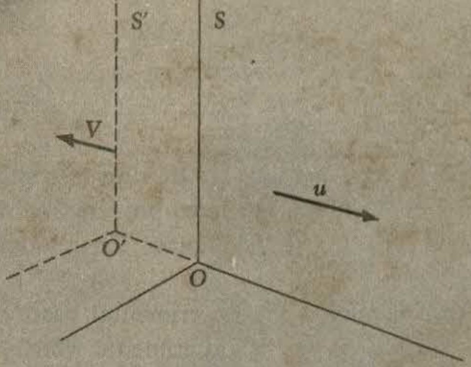
Speed of Light in Inertial Frames in Relative Motion

An elementary application of the Galilean transformation to the problem of a moving receiver requires that the speed of light

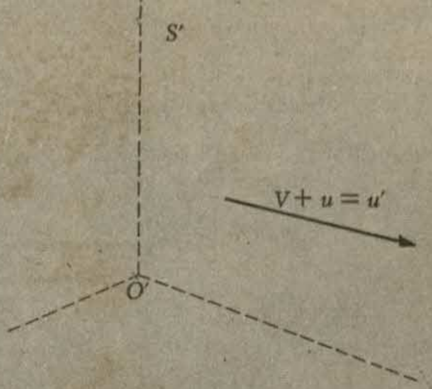
---

† Note a little computational trick: Suppose that $y = Ax^n$, where A, n are constants, and we want to find dy/y in terms of dx/x. We take the natural logarithm of both sides to form $\log y = \log A + n \log x$. We then take differentials of both sides to obtain $dy/y = n\,dx/x$. Here we have used the relation $d \log x/dx = 1/x$.

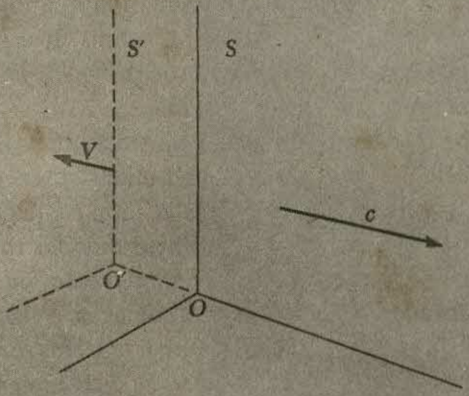
‡ See G. C. McVitties, *Physics Today*, p. 70 (July 1964).



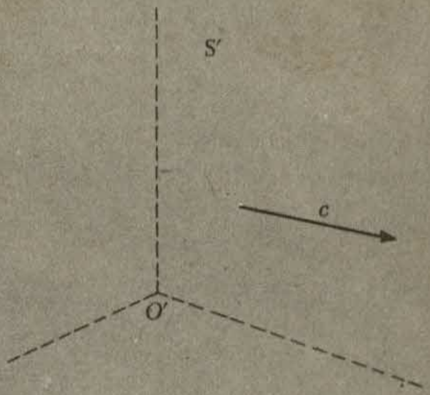
If u is an ordinary terrestrial speed as observed in inertial frame S, the Galilean transformation tells us that . . .



in inertial frame S′ we will observe $u' = V + u$.



However experiments show that if an object has speed c in S, . . .



it also has speed c in S′.

in the frame of the receiver be different from c. According to common sense we expect the speed of light c_R relative to the moving receiver to be given by

$$c_R = c \pm V, \tag{25}$$

where V is the speed of the receiver that is supposed to be moving toward (+) or away (−) from the source. This seems a perfectly reasonable way to add velocities. The same relation should hold when the source and receiver are at rest and the medium moves with velocity V. The relation (25) is apparently obeyed in countless everyday experiences, at least where light is not involved. It holds for sound waves, if v_s is written for c. But it is *not true*, even approximately, for light waves in free space. It is found experimentally that

$$c_R = c, \tag{26}$$

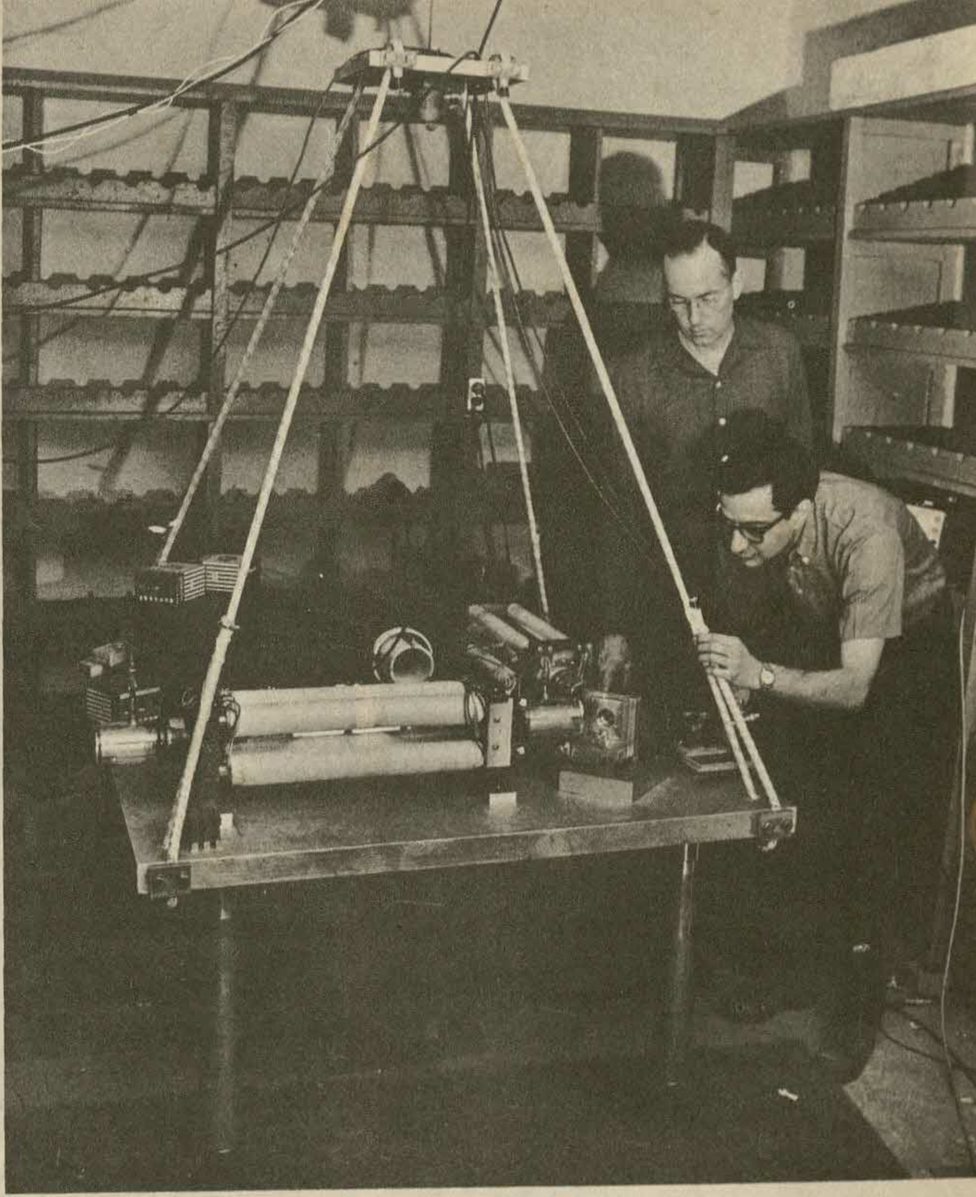
for any frame *regardless of its velocity*, and regardless of the velocity relative to an imagined propagation medium. This demonstrated fact lies at the root of the relativistic formulation of physical laws.

We now examine the experimental basis of (26). There are many different types of experiments which support the special theory of relativity; those leading to (26) make a convenient takeoff point. We consider the experiments which show that the velocity of light is independent of the velocity (3×10^6 cm/sec) of the earth in its orbit.

First suppose, as did the physicists of the nineteenth century, that light propagates as an oscillation in a medium, just as sound propagates as an oscillation of atoms in a liquid, solid, or gas. The luminiferous medium through which light waves propagate in free space was called the *ether*.

What is ether? Today we consider ether as only another word for vacuum. But Maxwell and many others could not imagine a field as a self-supporting institution propagating in free space. Maxwell argued: "But in all these theories the question naturally occurs:—If something is transmitted from one particle to another at a distance, what is its condition after it has left the one particle and before it has reached the other? If this something is the potential energy of the two particles, as in Neumann's theory, how are we to conceive this energy as existing in a point of space, coinciding neither with the one particle nor with the other? In fact, whenever energy is transmitted from one body to another in time, there must be a medium or substance in which the energy exists after it leaves one body and before it reaches the other, for energy, as Torricelli remarked, 'is a quintessence of so subtle a nature that it cannot be contained in any vessel except the inmost substance of material things.' Hence all these theories lead to the conception of a medium in which the propagation takes place, and if we admit this medium as an hypothesis, I think it ought to occupy a prominent place in our investigations, and that we ought to endeavour to construct a mental representation of all the details of its action, and this has been my constant aim in this treatise."

The obvious direct experiment to test the possible dependence of the velocity of light on the motion of the earth is to time accurately the one-way passage of a pulse of light over a measured path. This would be done separately in both directions on a north-south line, and then on an east-west line, and finally over again after six months, when the velocity of the earth about the

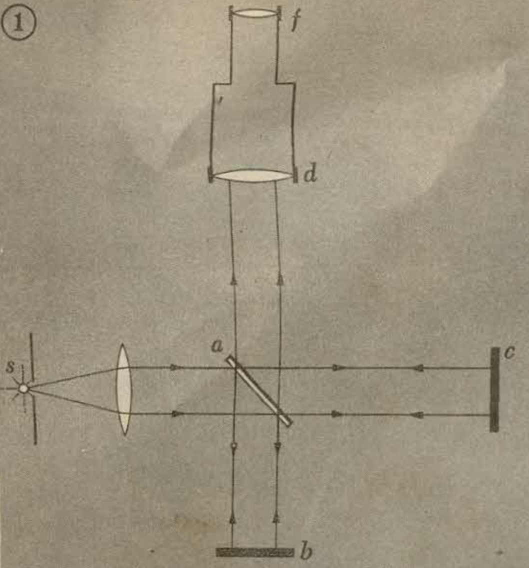


A precision apparatus for a relativistic optical experiment using two gas lasers. The site is a former wine cellar in Round Hill, Massachusetts. The workers are Charles H. Townes and Ali Javan.

sun has a reversed direction. With the development of lasers, sufficiently accurate clocks exist to permit such a direct experiment; the limiting technological factor at present appears to be the rise time of a pulse. At 10^{-9} sec this introduces an effective error of $10^{-9}\,c = 30$ cm in the length of the path. The clocks in such an experiment would have to be synchronized at one spot and then separated slowly to their final positions.

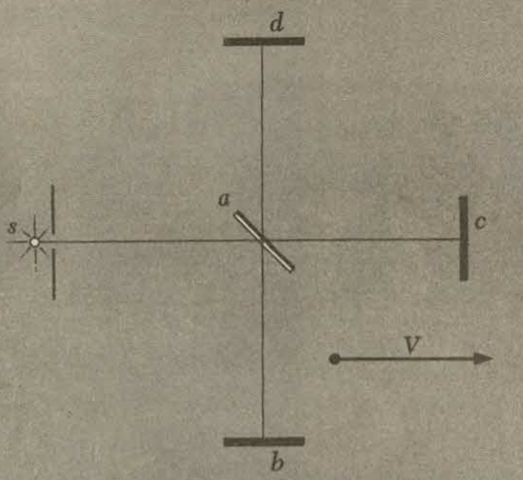
A number of experiments have been performed to test (25); that is, to detect ether shift. All have failed to show a movement of the earth through the ether; the most crucial were those carried out by Michelson and Morley.

①



The Michelson-Morley experimental interferometer consists of light source *s*, half-silvered mirror *a*, mirrors *b* and *c*, and telescope detector *d*: *f* represents focus of telescope. If interferometer is at rest in ether, an interference pattern between the beams *aba* and *aca* . . .

②

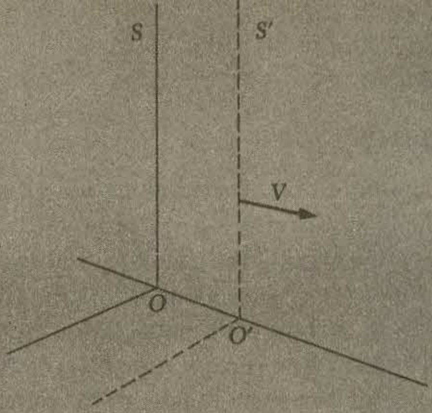


is observed at *d*. If apparatus (and earth) have velocity *V* with respect to hypothetical ether, we would expect interference pattern to change at *d*, since the times to traverse *aba*, *aca* would now change by different amounts. To see this . . .

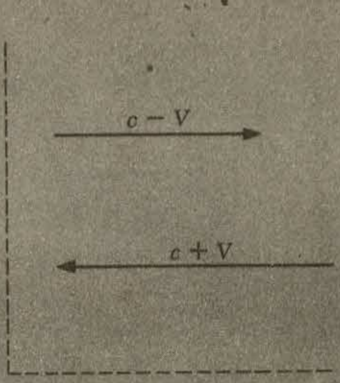
Michelson-Morley Experiments. Two sets of light waves derived from a common monochromatic source may interfere constructively or destructively at a point, according to the relative phase of the waves at that point. The relative phase may be changed by requiring one wave train to travel farther than the other. Michelson and Morley constructed an elaborate interferometer, the essential parts of which are contained in the figures. A beam of light from a single source *s* was split by a half-silvered mirror at *a*. We continue the description of the experiment in essentially the words and notation of the original workers:† "Let *sa* [Fig. 1] be a ray of light which is partly reflected in *ab*, and partly transmitted in *ac*, being returned by the mirrors *b* and *c*, along *ba* and *ca*. *ba* is partly transmitted along *ad*, and *ca* is partly reflected along *ad*. If then the paths *ab* and *ac* are equal, the two rays interfere along *ad*. Suppose now, the ether being at rest, that the whole apparatus moves in the direction *sc* with the velocity of the earth in its orbit; the directions and distances traversed by the rays will be altered

---

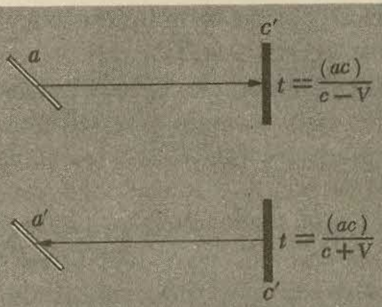
† A. A. Michelson and E. W. Morley, *Am. J. Sci.* 34, 333 (1887). This was one of the most remarkable experiments of the nineteenth century. Simple in principle, the experiment led to a scientific revolution with far-reaching consequences. Note that the ratio of the speed of the earth in its orbit to the speed of light is about 10^{-4}. In reproducing the excerpt, we have written *c* for their *V*, and *V* for their *v*; interpolated remarks are enclosed in brackets.



consider a Galilean frame S′ moving with earth and interferometer. S is a Galilean frame at rest in ether.



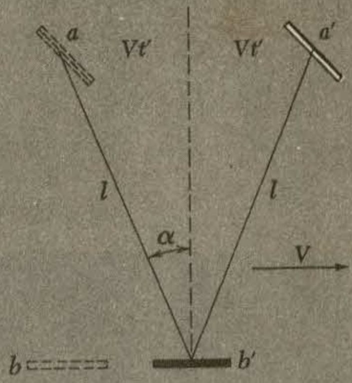
According to the Galilean transformation, light moving to right has speed $c - V$ in S′; light moving to left has speed $c + V$ in S′.



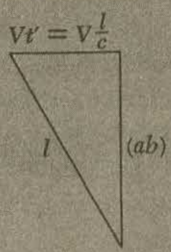
Thus the time to go from a to c' and back to a' is

$$\Delta t(ac'a') = \frac{(ac')}{c - V} + \frac{(ac')}{c + V},$$

where (ac') denotes the distance between a and c'.



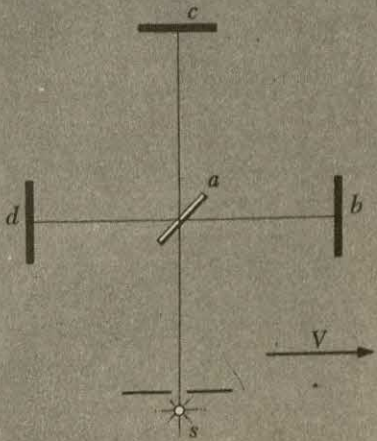
What is the time $\Delta t(ab'a') = 2t'$ to go from a to b and back to a? In the Galilean frame S at rest in ether, the interferometer has velocity V to right; light has speed c.



$$l^2 = \frac{V^2}{c^2}l^2 + (ab)^2$$

$$\therefore\ l = \frac{(ab)}{(1 - V^2/c^2)^{1/2}}$$

$\Delta t(ab'a') = 2t' = 2(ab)/\sqrt{c^2 - V^2}.$
Thus even if $(ab) = (ac)$, the Galilean transformation leads us to expect a . . .



shift in the interference pattern if the interferometer changes its velocity with respect to the ether. None was observed.

thus:—The ray sa is reflected along ab' [Fig. 2]; the angle $ab'a'$ being equal to the aberration α, is returned along $b'a'$, ($ab'a' = 2\alpha$) and goes to the focus of the telescope, whose direction is unaltered. The transmitted ray goes along ac', is returned along $c'a'$, and is reflected at a', making $c'a'd'$ equal $90° - \alpha$, and therefore still coinciding with the first ray. It may be remarked that the rays $b'a'$ and $c'a'$ do not now meet exactly in the same point a', though the difference is of the second order; this does not affect the validity of the reasoning. Let it now be required to find the difference in the two paths $ab'a'$ and $ac'a'$.

"Let c = velocity of light.
V = velocity of the earth in its orbit.
D = distance ab or ac [Fig. 1].
T = time light occupies to pass from a to c'.
T' = time light occupies to return from c' to a' [Fig. 2].

Then

$$T = \frac{D}{c - V}, \qquad T' = \frac{D}{c + V}. \tag{27a}$$

The whole time of going and coming is

$$T + T' = 2D\frac{c}{c^2 - V^2}, \tag{27b}$$

and the distance traveled in this time is

$$2D\frac{c^2}{c^2 - V^2} \cong 2D\left(1 + \frac{V^2}{c^2}\right), \tag{27c}$$

neglecting terms of the fourth order. The length of the other path is evidently

$$2D\sqrt{1 + \frac{V^2}{c^2}}, \tag{27d}$$

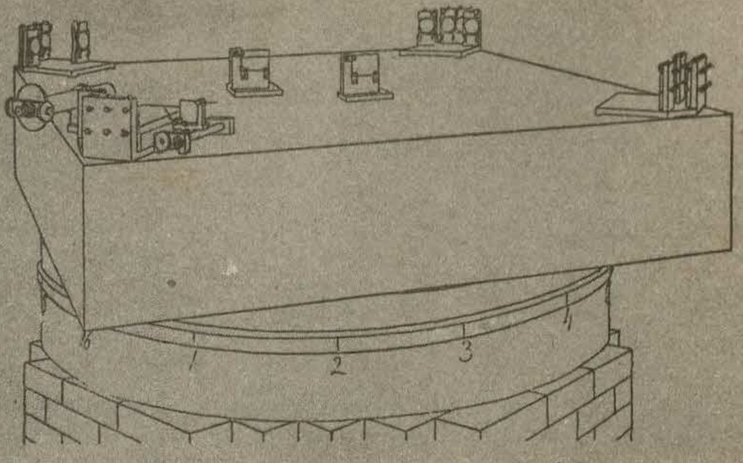
or to the same degree of accuracy,

$$2D\left(1 + \frac{V^2}{2c^2}\right). \tag{27e}$$

The difference is therefore

$$D\frac{V^2}{c^2}. \tag{27f}$$

If now the whole apparatus be turned through 90°, the difference



Perspective of the apparatus described by Michelson and Morley in their 1887 paper.

will be in the opposite direction, hence the displacement of the interference fringes should be $2D(V^2/c^2)$. Considering only the velocity of the earth in its orbit, this would be $2D \times 10^{-8}$. If, as was the case in the first experiment, $D = 2 \times 10^6$ waves of yellow light, the displacement to be expected would be 0.04 of the distance between the interference fringes.

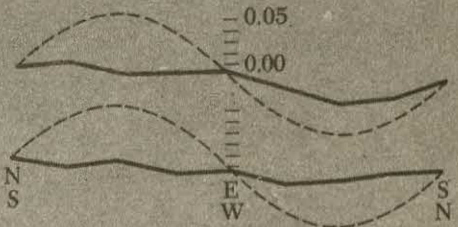
"In the first experiment one of the principal difficulties encountered was that of revolving the apparatus without producing distortion; and another was its extreme sensitiveness to vibration. This was so great that it was impossible to see the interference fringes except at brief intervals when working in the city, even at two o'clock in the morning. Finally, as before remarked, the quantity to be observed, namely, a displacement of something less than a twentieth of the distance between the interference fringes may have been too small to be detected when masked by experimental errors.

"The first named difficulties were entirely overcome [in the second experiment] by mounting the apparatus on a massive stone floating on mercury; and the second by increasing, by repeated reflection, the path of the light to about ten times its former value.

". . . Considering the motion of the earth in its orbit only, this displacement should be

$$2D\frac{V^2}{c^2} = 2D \times 10^{-8}.$$

The distance D was about eleven meters, or 2×10^7 wavelengths of yellow light; hence the displacement to be expected was 0.4 fringe [if the earth were traveling through an ether]. The actual displacement was certainly less than the twentieth part of this,



"The results of the observations are expressed graphically [in the figure]. The upper is the curve for the observations at noon, and the lower that for the evening observations. The dotted curves represent *one-eighth* of the theoretical displacements. It seems fair to conclude from the figure that if there is any displacement due to the relative motion of the earth and the luminiferous ether, this cannot be much greater than 0.01 of the distance between the fringes." (Michelson and Morley, *ibid.*) The vertical axis is the displacement of the fringes; the horizontal axis refers to the orientation of the interferometer relative to an east-west line.

and probably less than the fortieth part. But since the displacement is proportional to the square of the velocity, the relative velocity of the earth and the ether is probably less than one-sixth the earth's orbital velocity, and certainly less than one-fourth."

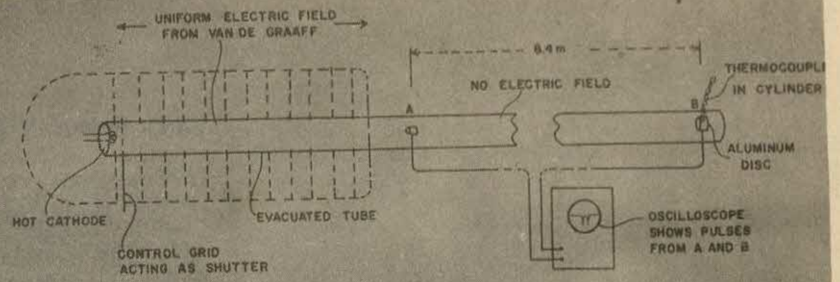
The experimental results of Michelson and Morley were contrary to what we would expect, based on the Galilean transformation. The experiments have since been repeated (with variations) with different wavelengths of light, with starlight, with extremely monochromatic light from a modern laser, at high altitudes, under the earth's surface, on different continents, and at different seasons, over a period of some 80 years. We can say that the ether drift is zero to a precision which is best expressed by saying that the speeds of light upstream and downstream are equal within a variation of less than 10^3 cm/sec, or of 1 part in 1000 of the earth's orbital velocity about the sun.

Invariance of c. The null result of the Michelson-Morley experiment suggests that the effects of the ether are undetectable. This means that in considering Doppler effects in light propagation, the result should involve only the relative motion of two frames and not the absolute velocity relative to some fixed ether.† The result also suggests that the speed of light is independent of the motion of the source or of the observer. The experimental evidence on the latter point is quite good, but could be improved. The work by Sadeh quoted in Chap. 11 shows that the velocity of γ rays is constant within ± 10 percent, independent of the velocity of the source, for source velocities of the order of $\frac{1}{2}c$. We conclude from all the experimental evidence that *a spherical wavefront of light emitted from a point source in one inertial frame will appear as spherical to an observer in any other inertial frame.*

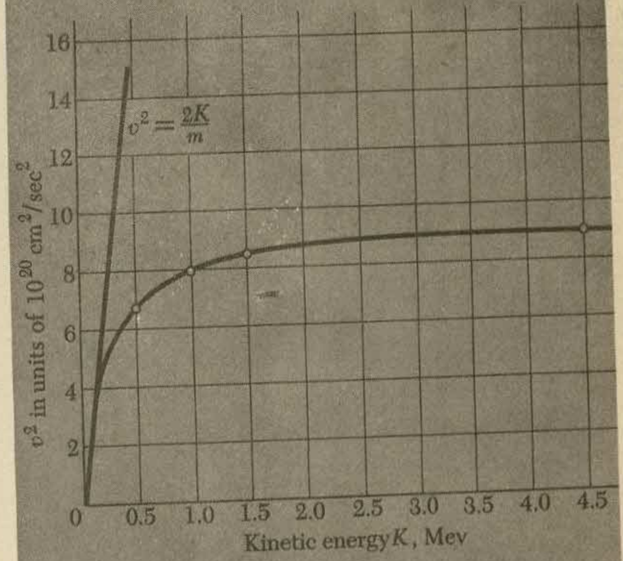
We noted in an earlier section that the speed of electromagnetic waves is independent of frequency over the range 10^8 to 10^{22} cps. Careful measurements also show that c is independent of the intensity of the light and also of the presence of other electric and magnetic fields. Our discussions have been limited entirely to electromagnetic waves traveling in free space.

---

† Note that in this respect light propagation is unlike sound propagation. In treating the Doppler effect of sound we had to know the velocity of the medium relative to transmitter and receiver. The Michelson-Morley experiment tells us that for light propagation in free space we must forget about the ether.


UNIFORM ELECTRIC FIELD FROM VAN DE GRAAFF
8.4 m
NO ELECTRIC FIELD
THERMOCOUPL
IN CYLINDER
A
B
ALUMINUM DISC
HOT CATHODE
EVACUATED TUBE
CONTROL GRID ACTING AS SHUTTER
OSCILLOSCOPE SHOWS PULSES FROM A AND B

The general arrangement of the ultimate speed experiment. The electrons are accelerated in a uniform field on the left and timed between *A* and *B* by the oscilloscope.



The Ultimate Speed. We have seen that electromagnetic waves in free space can only travel with the speed c. Can the speed of anything exceed the speed limit c?

Consider the motion of charged particles in an accelerator. Can particles be accelerated to travel faster than c? We have not as yet in this course encountered directly any principle which prevents the acceleration of charged particles to arbitrarily high velocities.

The following experiment† illustrates the proposition that a particle cannot be accelerated to a speed greater than c. Electrons are accelerated by successively larger electrostatic fields in a Van de Graaff accelerator, after which the electrons drift with constant velocity through a field-free region. Their time of flight, and hence their velocity over a measured distance AB, is measured directly, and the kinetic energy (which is turned to heat at the target at the end of the path) is measured by means of a calibrated thermocouple.

In the experiment the accelerating potential φ is known with good precision. The kinetic energy of an electron is

$$K = eEL = e\varphi,$$

where L is the distance over which acceleration occurs and $\varphi = EL$ is the difference in electric potential between the ends of the accelerating path. If $\varphi = 10^6$ volts, the electron after acceleration has an energy of 1×10^6 electron volts (1 Mev). Now 10^6 volts $\cong 10^6/300$ statvolts, so that the kinetic energy acquired by an electron is

$$\frac{(4.80 \times 10^{-10})(10^6)}{300} \cong 1.60 \times 10^{-6} \text{ erg.} \tag{28}$$

---

† This experiment was performed by W. Bertozzi in connection with the PSSC film "The Ultimate Speed." Our account draws directly from chap. A-3 of the PSSC Advanced Topics Program. See *Am. J. Phys.* **32**, 551 (1964).

If N electrons per second travel in the beam, the power delivered to the aluminum target at the end of the beam should be $1.60 \times 10^{-6} N$ erg/sec. This agrees exactly with the direct thermocouple determination of the power absorbed by the target. This result confirms that the electrons deliver to the target the kinetic energy acquired during their acceleration. Further, on the basis of nonrelativistic mechanics, we expect that

$$K = \tfrac{1}{2}mv^2, \tag{29}$$

so that a graph of v^2 against the kinetic energy K should be a straight line. For energies greater than about 10^5 ev, however, the linear relation between v^2 and K does *not* hold experimentally. Instead, the velocity is observed to approach the limiting value 3×10^{10} cm/sec at higher energies. The experimental results may be summarized: The electrons absorb the expected energy from the accelerating field, but their velocity does not increase without limit. Many other experiments suggest, as this one does, that c is the upper limit to the velocity of particles. Thus we believe firmly that c is the maximum signaling speed with either particles or electromagnetic waves: *c is the ultimate speed.*

We are now prepared to study special relativity in Chap. 11 with the knowledge from experiment that:

1. c is invariant among inertial frames.
2. c is the maximum signaling speed.
3. Only relative velocities of inertial frames are significant for light propagation.
4. Galilean invariance is *inadequate* at high relative velocities, because measurements of length and kinetic energy in inertial systems must represent the results 1 and 2.
5. Because v^2 at high velocities is not proportional to the kinetic energy, we may expect to find that the inertial mass varies with the velocity.

We have reviewed only a very small fraction of the experiments that support the special theory of relativity, which is now very solidly established. Physicists place as much reliance on this theory as any other part of physics. Our next endeavor must be to formulate it precisely and to understand some of its major consequences.

Film list

"Speed of Light" (21 min) W. Siebert (PSSC-MLA). The velocity of light is measured by the time of flight of a light pulse and also by the rotating-mirror method.

"Doppler Effect of Waves" (4 min) (ESI film). A beautiful demonstration. It is part of a set of nine demonstrations of wave phenomena photographed in a ripple tank.

"The Ultimate Speed" (37½ min) W. Bertozzi (ESI film). The relationship between the kinetic energy of electrons and their speed; investigated by time-of-flight and calorimetric techniques. The results indicate a limiting speed equal to c in agreement with the special theory of relativity.

Further reading

A. A. Michelson, *Studies in optics* (University of Chicago Press, Chicago, 1927; paperback reprint, 1962).

Problems

1. *Doppler shift.* A space navigator wishes to determine his velocity of approach as he nears the moon. He sends a radio signal of frequency $\nu = 5000$ Mc/sec and compares this frequency with its echo, observing a difference of 86 kc/sec. Calculate the velocity of the space vehicle relative to the moon. (The nonrelativistic expression for the Doppler effect is sufficiently accurate for many purposes.) *Ans.* 2.6×10^5 cm/sec.

2. *Recessional red shift.* A spectral line appearing at a wavelength of 5000 Å in the laboratory is observed at 5200 Å in the spectrum of light coming from a distant galaxy.

(*a*) What is the recessional velocity of the galaxy? *Ans.* 1.2×10^9 cm/sec.

(*b*) How far away is the galaxy? *Ans.* 4×10^{26} cm.

3. *Speed of light.* In Michelson's celebrated measurement of the speed of light, an octagonal reflecting prism rotating about the axis of the prism reflected a beam of light from a distant light source and back to an observer near the source. The timing provides that the transit time of the light equal one-eighth of the period of rotation of the octagonal prism. The one-way distance was $L = 35.410 \pm 0.003$ km and the frequency of rotation of the prism was $\nu = 529$ cycles per sec, to an accuracy of 3×10^{-5} cycles per sec.

(*a*) Calculate the speed of light from these data. (A fractional correction of the order of 10^{-5} for atmospheric effects had to be applied.)

(*b*) The angle between any two adjacent prism faces was $135° \pm 0.1''$. Estimate the overall precision of the measurement of c.

4. *Eclipses of Io.* One of Jupiter's satellites, Io, moves in an orbit of radius 4.21×10^{10} cm with an average period of 42.5 hours. Roemer observed that the period varied regularly during the year, with a period of variation of about one year. The maximum deviation of the period from

the average was 15 seconds, at times approximately six months apart. Neglect the orbital travel of Jupiter.

(*a*) Estimate the distance the earth travels in one period of Io's motion about Jupiter. *Ans.* 4.5×10^{11} cm.

(*b*) When does Io's period appear to be greatest?

(*c*) Use the preceding result and the data provided to estimate the velocity of light.

(*d*) Estimate the accumulated delay in the six months following the point of zero delay when the Earth is closest to Jupiter.

(*e*) Consider the Io-Jupiter system as a light source modulated at the frequency (1/42.5) cycles per hour. Treat the variation in observed period as a Doppler effect, and calculate the speed of light accordingly.

5. *Stellar parallax.* Stellar parallax was predicted by Aristarchus of Samos (ca. 200 B.C.) and it was finally observed for certain by Bessel in 1838. A notably unsuccessful attempt was made by Bradley, who discovered instead the aberration of starlight. During the course of a year the apparent position of a star shifts between extremes by approximately 40″ of arc, due to aberration.

(*a*) What would be the distance in parsecs of a star with a parallax of 20″? The nearest known star is α Centauri at a distance of about 1.3 parsecs. *Ans.* 0.05 parsecs.

(*b*) Show that the apparent annual motion from aberration of stars near the ecliptic is a straight line whose ends subtend a 40″ angle.

6. *Rotation of galaxies.* In 1916, before the great distances of the nebulae (galaxies) were known, the spiral M101 was reported to rotate like a solid body with a period of 85,000 years. The observed angular diameter is 22′. Calculate the maximum possible distance of the nebula if the above period is correct, supposing that the extremities of the nebula are not to move faster than c. [Recent measurements of stars in M101 place it at a distance of 8.5×10^{24} cm. It is apparent that the rotation reported in 1916 was overestimated.]

7. *Variable stars.* The 200-in. Mt. Palomar telescope can barely resolve individual stars in galaxies at a distance of 3×10^{25} cm. One method for calibrating distances of this order of magnitude involves observation of the periods in the luminosity of certain Cepheid-type variable stars. A Cepheid-type star is a gravitationally unstable star that exhibits periodic pulsations in which its radius may change by perhaps 5 to 10 percent. The temperature of the star changes with the same period as the radius, so that one observes periodic variations in brightness. Periods as short as a few hours have been found. A Cepheid whose intrinsic luminosity is 2×10^4 times that of the sun has a period of 50 days in our galaxy.

(*a*) Estimate from the distance-velocity relation the radial velocity for a galaxy at a distance of 3×10^{25} cm.

(*b*) What would we expect to observe for the period of this Cepheid in a galaxy at the distance cited above?

8. *Novae.* Occasionally a star is seen to experience an explosion in which a portion of its outer layers is thrown out with high velocity. Such a star is called a *nova.* A recent nova was observed visually to have a surrounding shell after its outburst. The angular diameter of the shell was found to increase by 0.3″ per year. The spectrum of the nova is a normal stellar spectrum with superimposed broad emission lines, the widths (in wavelengths) of which remain constant at 10 Å (in the vicinity of a wavelength of 5000 Å), though the lines are dimming. The width is to be interpreted as a measure of the Doppler shift between the parts of the shell advancing toward us and receding from us. Estimate the distance to the nova, if the shell is optically thin (so that we receive as much light from the far hemisphere as from the near). *Ans.* 1.2×10^{21} cm.

9. *Velocities of galaxies.* Measured radial velocities of galaxies relative to the Earth are not isotropic over the sky. Nonisotropy results from the motion of the sun (orbital velocity) with respect to the center of our galaxy, and from our galaxy's own motion with respect to the local extragalactic standard of rest. Let us examine all galaxies at a particular distance, say, 3.26×10^7 light-years.

(*a*) What is the mean radial velocity of these galaxies?

(*b*) Where in their spectra will be the average location of the Hα line of hydrogen? (In the laboratory, $\lambda_{H\alpha} = 6.563 \times 10^{-5}$ cm.)

In our sample we find that in a certain direction the velocities are 300 km/sec larger than the average and in just the opposite direction they are this much too small.

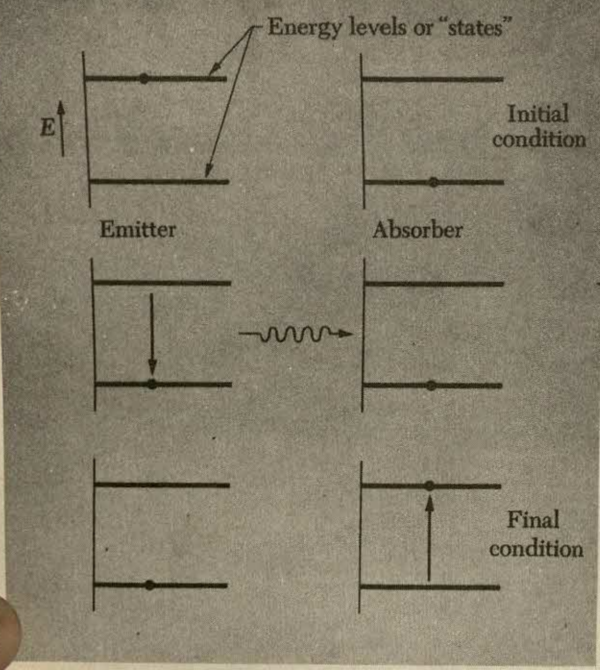
(*c*) What is the velocity of the sun in this frame of reference?

(*d*) Is that necessarily the orbital velocity of the sun around the center of our galaxy?

(*e*) Assuming that this is the orbital velocity, estimate the mass of our galaxy, taking all the mass to be at its center and the orbit of the sun to be circular (the distance to the center of the galaxy is 3500 light-years). Compare with the mass of 8×10^{44} gm quoted for the mass of the galaxy and explain the difference.

Ans. (*a*) The mean velocity of the galaxies as calculated from the velocity-distance relation is 930 km/sec. (*b*) The Hα line will be, on the average, at 6.584×10^{-5} cm. (*c*) 300 km/sec. (*d*) No, for it can include any motion of our galaxy as a whole in this reference frame. (*e*) 4.5×10^{43} gm. This is less than that usually quoted because much of the mass of our galaxy is not at the center—in fact, much mass lies exterior to the sun, where it would not affect the sun's motion or be detectable in this way.

10. *Rotation of stars.* The sun is seen from its surface features to rotate slowly, with a period of 25 days at the equator. Some stars, however, rotate far faster. How can this be determined in view of the fact that the stars are too distant to be seen except as points of light?



Advanced Topic. Recoilless emission of gamma rays

A nucleus in an excited energy state may emit a photon (gamma ray) in making a transition to the ground, or unexcited, state of the nucleus. The inverse process also may occur: A nucleus in its ground state may absorb a photon, leaving the nucleus in an excited state.

Suppose we prepare a source containing excited nuclei. In the course of time the source will emit photons. We allow the photons to strike an absorber which contains similar nuclei in their ground state. These nuclei will absorb the incident photons and will then reemit photons. The phenomenon of absorption and reemission is known as nuclear fluorescence. The photons emitted (by both the source and the absorber) will have a range of energy of approximate width Γ, as shown in the sketch. Here Γ is the Greek character capital gamma.

A good example is the nucleus Fe^{57}. This is formed in an excited state as the product of the radioactive decay of Co^{57}. The excited state of Fe^{57} emits a photon of energy 14.4 kev, leaving the Fe^{57} nucleus in its ground state.

Consider an Fe^{57} nucleus in an excited state, and suppose that the nucleus is initially at rest in free space. When the photon is emitted, the nucleus will recoil in the direction opposite to the photon.

(*a*) What is the frequency ν of a photon of energy 14.4 kev? Recall that $E = h\nu$, where h is Planck's constant and E is the energy.

Ans. 3.5×10^{18} cps.

(*b*) The momentum of the photon is $h\nu/c$. What is the recoil momentum of the nucleus? *Ans.* 7.7×10^{-19} gm-cm/sec.

(*c*) Show that the recoil energy R of the nucleus is

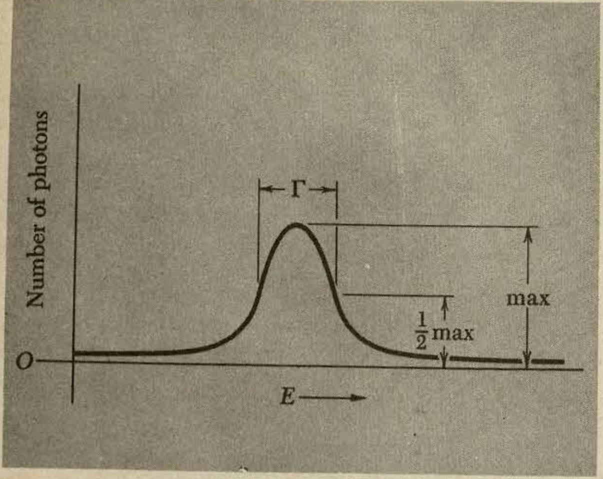
$$R = \frac{E^2}{2Mc^2},$$

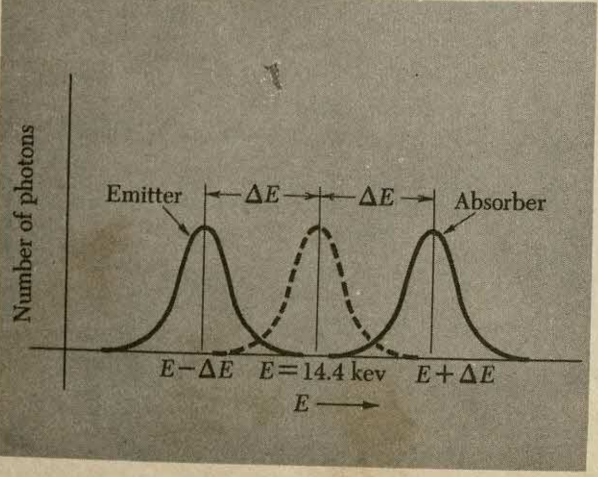
where M is the mass of the nucleus and E is the energy of the photon. Evaluate R in electron volts for Fe^{57}. *Ans.* 2×10^{-3} ev.

Nuclear energy levels are not perfectly sharp but have a width Γ according to the uncertainty principle,

$$\Gamma\tau \approx \frac{h}{2\pi},$$

where τ is the mean life of the state. For low-energy γ rays like those from Fe^{57}, the spread in energy of the nuclear energy levels may be much less than the recoil energy R. In this situation the emitted γ ray cannot normally be reabsorbed by a nucleus in the ground state, because the frequency is not right.





One method of bringing emitter and absorber frequencies effectively into tune is to give the source a velocity relative to the absorber.

(*d*) What is the required magnitude of this velocity for Fe^{57}?

(*e*) Mössbauer observed that in some of the emissions from certain crystals, the recoil momentum is taken by the crystal as a whole rather than by the individual nucleus. At room temperature, about 70 percent of the photons from an Fe crystal are almost recoilless in this sense. Calculate R for a recoilless photon if the mass of the crystal of Fe is 1 gm.

Ans. 2×10^{-25} ev, which is entirely negligible.

Chapter 11

Lorentz transformation 346
Examples: Aberration of light 350
Addition of velocities 350
Measurement of length perpendicular to relative velocity 353
Length contraction 354
Time dilation of moving clocks 356
Examples: Lifetime of π^+ mesons 358
Longitudinal Doppler effect 361
Accelerated clocks 363
Further reading 363
Film list 364
Problems 364
Mathematical Note: Space-time 366
Historical Note: Simultaneity in special relativity 375

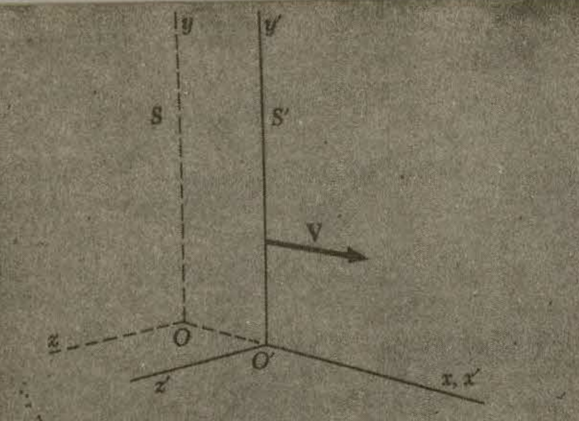
Lorentz Transformations of Length and Time

The null result of the Michelson-Morley experiment to detect the drift of the earth through the ether can only be understood by making a revolutionary change in our thinking; the new principle we need is simple and clear:

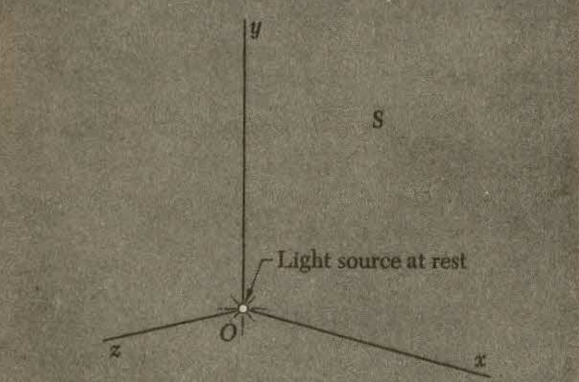
> The speed of light is independent of the motion of the light source or receiver.

That is, the speed of light is the same in all reference frames in uniform motion with respect to the source. All the vast consequences of the special theory of relativity follow from this single new assumption, which is to be added to our earlier assumptions that space is isotropic and uniform, and that the fundamental laws of physics are identical for any two observers in uniform relative motion.

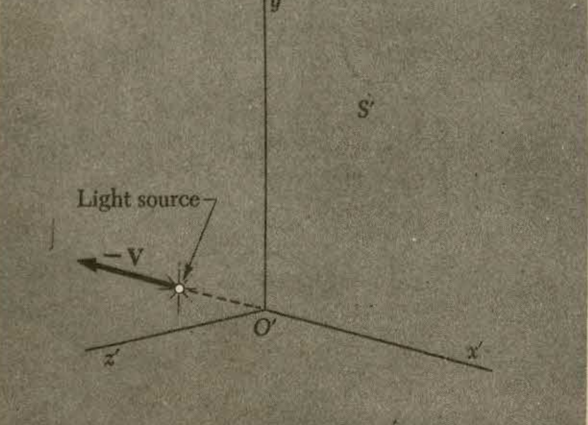
Electromagnetic waves or photons are not unique in having a velocity independent of the motion of the source. We believe that any particle with zero rest mass will have a velocity equal to *c* and independent of the motion of the source; particularly is this true of neutrinos and antineutrinos. We shall discuss photons, however, because it is easier to carry out experiments with photons than with neutrinos.



S and S′ are two inertial frames. S′ has velocity **V** with respect to S.



Suppose a light source is at rest at the origin in S.



The source has velocity −**V** in S′.

Consider first a light wave spreading out from a point source. The wavefront (surface of equal phase) will be a sphere if viewed in the reference frame in which the source is at rest. But according to our new principle the wavefront must also be a sphere when viewed in a reference frame in uniform motion with respect to the source; otherwise we could tell from the shape of the wavefront that the source is moving. The fundamental assumption that the speed of light is independent of the motion of the source demands that we be unable to tell from the shape of the wavefront whether or not the source is in uniform motion.

Lorentz Transformation. We now look for a transformation which will make the speed of light independent of the motion of the source or receiver. Let the reference frome S in which the source is at rest be the unprimed frame. Positions and times measured by an observer in this frame will be denoted by the unprimed symbols x, y, z, t. If a light source is at the origin of the frame S and a front is emitted at $t = 0$, the equation of a spherical wavefront is

$$x^2 + y^2 + z^2 = c^2t^2. \tag{1}$$

Equation (1) describes a spherical surface whose radius expands at the speed c.

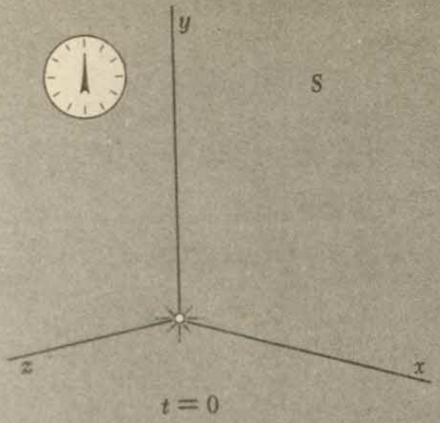
Let the moving reference frame S′ be the primed frame. Positions and times measured by an observer in this frame are denoted by the primed symbols x', y', z', t'. For convenience we suppose that the zero of t' coincides with the zero of t, and that the origin of x', y', z' coincides with the position of the light source in S at this zero of time. Then to an observer in S′ the equation of the spherical wavefront must be

$$x'^2 + y'^2 + z'^2 = c^2t'^2. \tag{2}$$

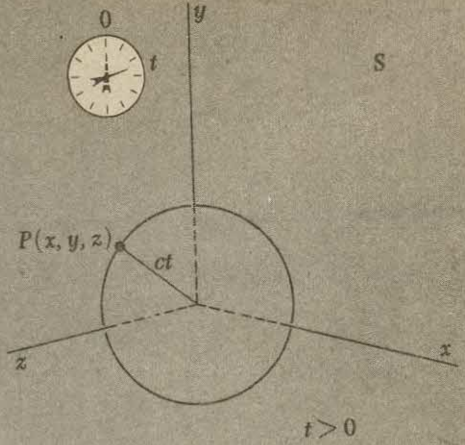
Here we have taken the same value c for the speed of light as in the frame S.

Suppose that the frame S′ is moving in the $+x$ direction with the constant velocity V with respect to S. The Galilean transformation (Chap. 3) connects measurements in the two frames according to the equations

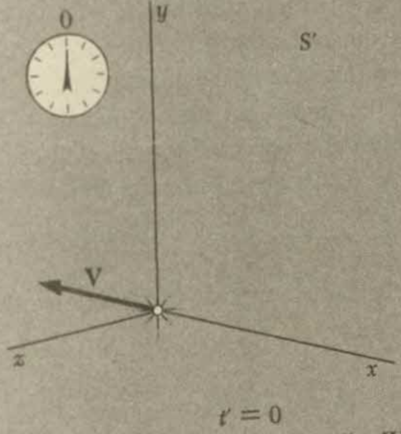
$$x' = x - Vt; \qquad y' = y; \qquad z' = z; \qquad t' = t. \tag{3}$$



Suppose the source emits a light pulse at time $t = 0$, as read by a clock at rest in S.

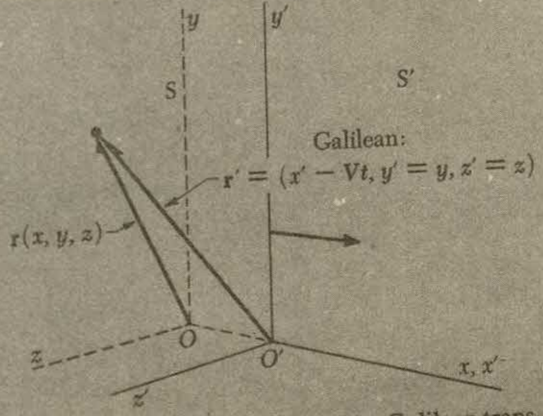


The spherical wavefront will reach point $P(x,y,z)$ at time $t = (x^2 + y^2 + z^2)^{1/2}/c$. The speed of the wavefront is c.



We can also observe the source from S′. We can set a clock which is at rest in S′ to read zero when the pulse is emitted.

The Galilean transformation predicts that in S′ we observe a spherical wavefront centered about the origin O of S. We actually observe a spherical wavefront in S′ centered about the origin O'. The same point P (now labeled with coordinates x', y', z') is reached at time $t' = (x'^2 + y'^2 + z'^2)^{1/2}/c$.



Obviously the trouble is that the Galilean transformation contradicts the postulates of special relativity. We need a new transformation. Perhaps we will find $t' \neq t$.

If we substitute (3) in (2) we obtain directly that

$$x^2 - 2xVt + V^2t^2 + y^2 + z^2 = c^2t^2, \tag{4}$$

which is certainly not in agreement with (1). Thus the Galilean transformation fails. If the principle of the constancy of the speed of light is valid, there should exist *some* transformation which reduces to the Galilean for $V/c \to 0$ and which transforms $x'^2 + y'^2 + z'^2 = c^2t'^2$ into $x^2 + y^2 + z^2 = c^2t^2$.

We suspect that the new transformation must be trivial for y' and z', because the terms in y'^2 and z'^2 in (2) transform into y^2 and z^2 in (1) without anything extra. We need a transformation which is *linear* in x and t, because we want to get a sphere which expands at a uniform rate. It is no use trying $x' = x^{1/2}t^{1/2}$ or $x' = \sin x$ or other functions. It is clear from (4) that we cannot leave the $t' = t$ transformation unchanged *if* we want to cancel out the undesired terms $-2xVt + V^2t^2$, because *something certainly must be added* to cancel these terms.

Let us try next a transformation of the form

$$x' = x - Vt; \qquad y' = y; \qquad z' = z; \qquad t' = t + fx, \tag{5}$$

where f is a constant to be determined. Then (2) becomes

$$x^2 - 2xVt + V^2t^2 + y^2 + z^2 = c^2t^2 + 2c^2ftx + c^2f^2x^2. \tag{6}$$

Notice that the terms xt cancel if we set

$$f = -\frac{V}{c^2}, \qquad \text{or} \qquad t' = t - \frac{Vx}{c^2}. \tag{7}$$

With this value of a, (6) can be written as

$$x^2\left(1 - \frac{V^2}{c^2}\right) + y^2 + z^2 = c^2t^2\left(1 - \frac{V^2}{c^2}\right). \tag{8}$$

This is closer to (1), but there remains an unwanted scale factor $(1 - V^2/c^2)$ multiplying x^2 and t^2.

We can dispose of the scale factor by taking the transformation to be

$$\boxed{\begin{aligned} x' &= \frac{x - Vt}{(1 - V^2/c^2)^{1/2}}; \qquad y' = y; \qquad z' = z; \\ t' &= \frac{t - (V/c^2)x}{(1 - V^2/c^2)^{1/2}}. \end{aligned}} \tag{9}$$

This is the *Lorentz transformation.*† It is linear in x and t; it reduces to the Galilean transformation for $V/c \to 0$; when substituted in (2) it gives

$$x^2 + y^2 + z^2 = c^2t^2, \tag{10}$$

exactly as required. That is,

$$x'^2 + y'^2 + z'^2 = c^2t'^2 \tag{11}$$

is *invariant* under a Lorentz transformation. The form of the equation describing the wavefront is the same in all frames moving with uniform relative velocity. Equation (9) is the unique solution to all our difficulties. The student should memorize the Lorentz transformation. It is no harder to learn by heart than any one tense of one irregular verb in a foreign language.

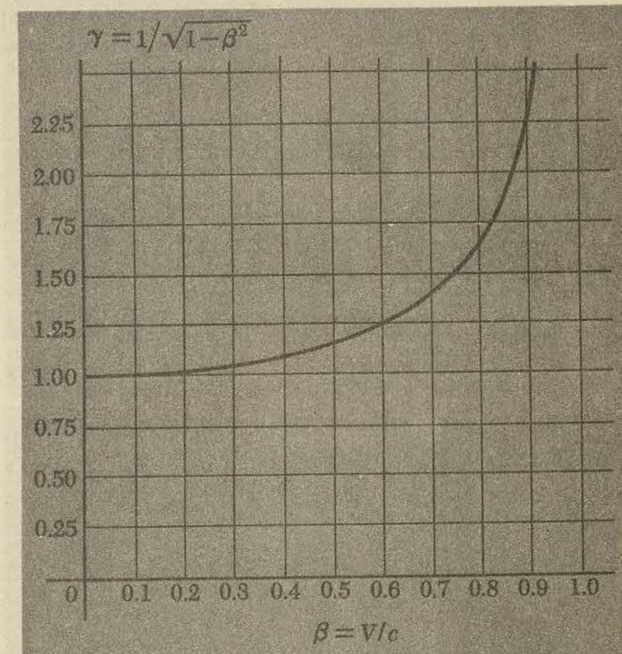
Now the propagation of electromagnetic waves is a consequence of the equations of electromagnetic theory. It is not surprising that the invariance of (11) has powerful implications for the form of the electromagnetic equations. We exploit these implications in Vol. II, where we develop the electromagnetic equations with the help of the Lorentz transformation (9).

It is sometimes convenient to make use of some standard notation. We introduce the notation

$$\boxed{\beta \equiv \frac{V}{c}.} \tag{12}$$

That is, β (Greek beta) is the velocity measured in a natural system of units in which $c = 1$. It is also convenient to introduce γ (Greek gamma):

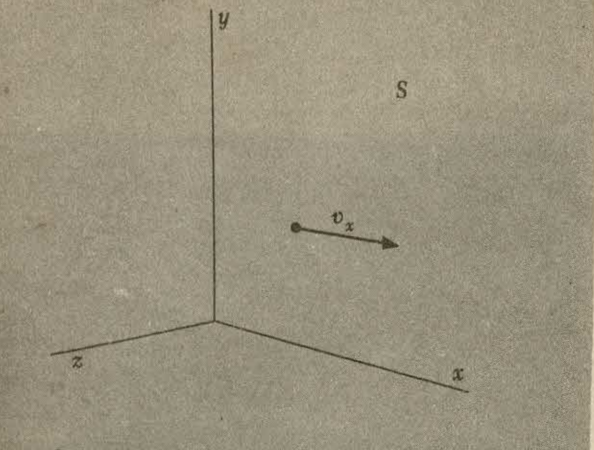
$$\boxed{\begin{aligned}\gamma &\equiv \frac{1}{(1-\beta^2)^{1/2}}\\ &\equiv \frac{1}{(1-V^2/c^2)^{1/2}}.\end{aligned}} \tag{13}$$



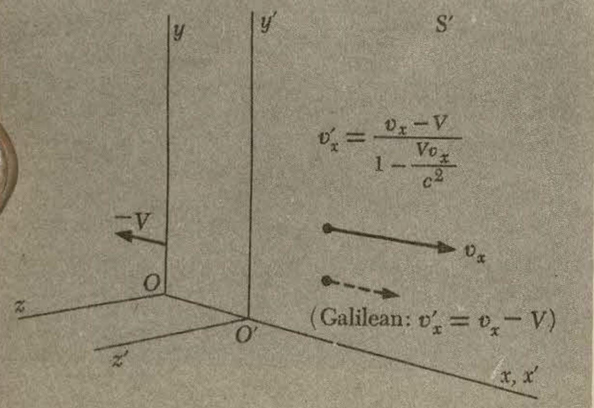
For $V \ll c$, $\gamma \simeq 1$. That is why the Galilean transformation is quite adequate for objects moving with ordinary speeds.

---

† This transformation has a long history. It was first used by J. Larmor to explain the null result of the Michelson-Morley experiment, in his *Aether and matter*, pp. 174–176 (Cambridge University Press, New York, 1900). Larmor claims accuracy only to order v^2/c^2; in fact, his results are exact.

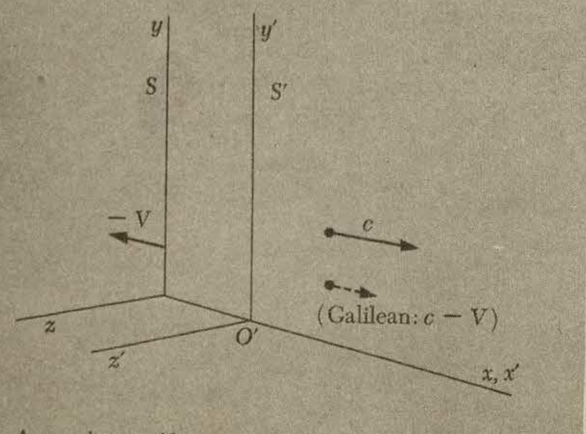


Suppose a particle has velocity v_x in S.



y
y'
S′
$v'_x = \frac{v_x - V}{1 - \frac{Vv_x}{c^2}}$
$-V$
v_x
O
O'
z
z'
(Galilean: $v'_x = v_x - V$)
x, x'

Then in S′ the Lorentz transformation predicts $v_x' = (v_x - V)/(1 - v_xV/c^2)$. The Galilean transformation would predict $v_x' = v_x - V$.



As we know, if $v_x = c$, $v_x' = c$ also, according to the Lorentz transformation. *This* was built into our theory from the beginning.

Note that $\gamma \geqq 1$. In extreme relativistic problems $1 - \beta \ll 1$; it is useful then to note that $1 - \beta^2 = (1 - \beta)(1 + \beta) \cong 2(1 - \beta)$. The Lorentz transformation (9) becomes

$$x' = \gamma(x - \beta ct); \qquad y' = y; \qquad z' = z; \qquad t' = \gamma\left(t - \frac{\beta x}{c}\right), \tag{14}$$

and the inverse transformation is easily seen to be

$$x = \gamma(x' + \beta ct'); \qquad y = y'; \qquad z = z'; \qquad t = \gamma\left(t' + \frac{\beta x'}{c}\right). \tag{15}$$

This is left to the reader as Prob. 2.

Example. Aberration of light. We saw in (10.3) that for a star directly overhead (when the earth's velocity v_e is perpendicular to the line of observation) the tilt angle, or aberration, of the telescope is given by

$$\tan\alpha = \frac{v_e}{c}. \tag{16}$$

This result was derived using a nonrelativistic argument. Now consider the problem relativistically, as an elementary exercise in the use of the Lorentz transformation.

Suppose that in the reference frame S in which the star is at rest a light signal from the star is observed along the z axis, with $x = y = 0$. The reference frame S′ in which the earth is at rest travels with velocity v_e in the x direction. Then the trajectory of the light signal is found directly from (14), with $x = 0$:

$$x' = -\gamma\beta ct; \qquad z' = z = ct; \qquad ct' = \gamma ct. \tag{17}$$

The tilt angle is then

$$\tan\alpha = \frac{(-x')}{z'} = \gamma\beta = \frac{v_e/c}{(1 - v_e^2/c^2)^{1/2}}. \tag{18}$$

The result (18) agrees with the nonrelativistic result (16) within the accuracy of measurement, because $v_e/c \cong 10^{-4}$ for the earth. The correct result is (18).

Example. Addition of velocities. Suppose the S′ reference frame moves with uniform velocity $V\hat{\mathbf{x}}$ relative to the S reference frame. A particle moves with uniform velocity components

v_x', v_y', v_z' relative to the S′ frame. What are the velocity components v_x, v_y, v_z of the particle relative to the S frame?

From (15) we have, with $\beta = V/c$,

$$x = \gamma x' + \gamma\beta c t'; \qquad t = \gamma t' + \frac{\gamma\beta x'}{c}, \tag{19}$$

whence

$$dx = \gamma\, dx' + \gamma\beta c\, dt'; \qquad dt = \gamma\, dt' + \frac{\gamma\beta\, dx'}{c}. \tag{19a}$$

Thus

$$v_x = \frac{dx}{dt} = \frac{\gamma\, dx' + \gamma\beta c\, dt'}{\gamma\, dt' + \gamma\beta\, dx'/c} = \frac{v_x' + \beta c}{1 + v_x'\beta/c}, \tag{20}$$

or

$$\boxed{v_x = \frac{v_x' + V}{1 + v_x'V/c^2}.} \tag{20a}$$

This result may be compared with the Galilean result $v_x = v_x' + V$, as in Chap. 3. Similarly, because $y = y'$ and $z = z'$,

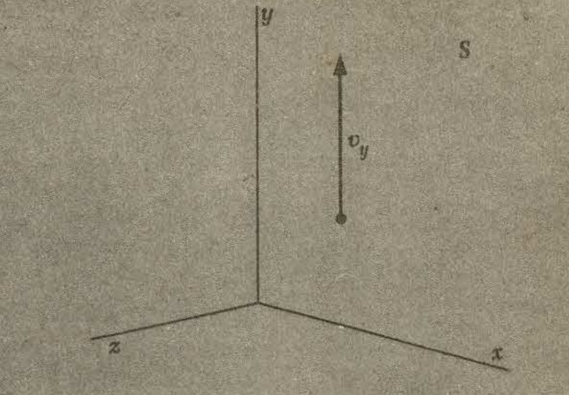
$$\boxed{\begin{aligned} v_y = \frac{dy}{dt} &= \frac{dy'}{\gamma\, dt' + \gamma\beta\, dx'/c} \\ &= \frac{v_y'}{1 + v_x'V/c^2}\left(1 - \frac{V^2}{c^2}\right)^{1/2}, \end{aligned}} \tag{20b}$$

and

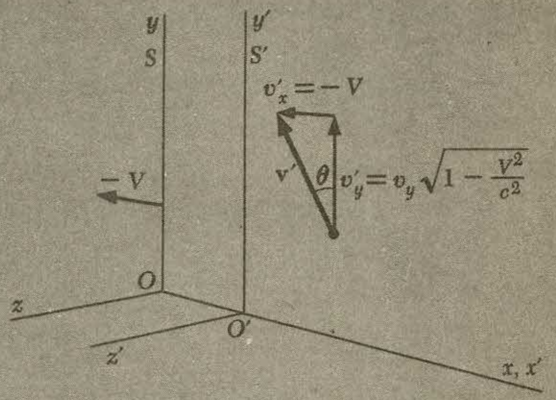
$$\boxed{v_z = \frac{v_z'}{1 + v_x'V/c^2}\left(1 - \frac{V^2}{c^2}\right)^{1/2}.} \tag{20c}$$

The inverse transformations follow from (14), or by solving (20*a*), (20*b*), (20*c*) for the primed velocity components:

$$\boxed{\begin{aligned} v_x' &= \frac{v_x - V}{1 - v_xV/c^2}; \\ v_y' &= \frac{v_y}{1 - v_xV/c^2}\left(1 - \frac{V^2}{c^2}\right)^{1/2}; \\ v_z' &= \frac{v_z}{1 - v_xV/c^2}\left(1 - \frac{V^2}{c^2}\right)^{1/2}. \end{aligned}} \tag{20d}$$

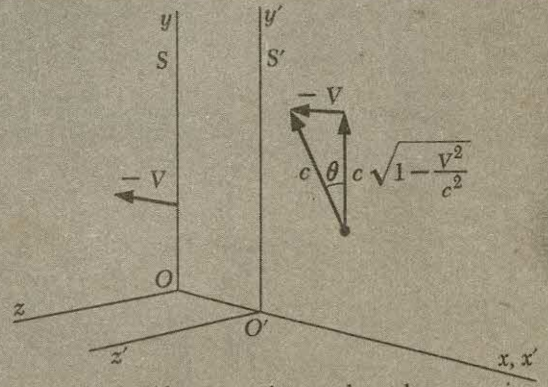


If a particle has velocity v_y in the y direction in S, . . .



it has the components shown in S′, according to the Lorentz transformation.

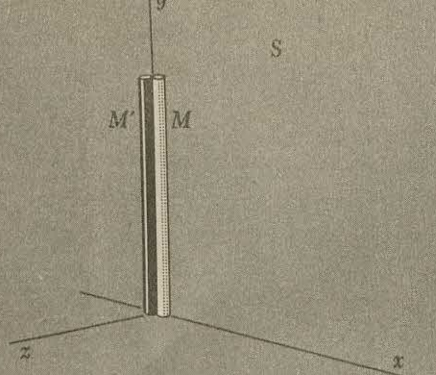
$$|\tan\theta| = \frac{V}{v_y\sqrt{1 - V^2/c^2}}.$$



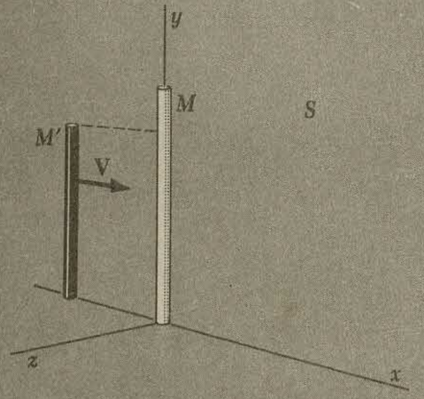
In particular if $v_y = c$, the resultant has magnitude c in S′. Thus

$$|\tan\theta| = \frac{V}{c\sqrt{1 - v^2/c^2}}.$$

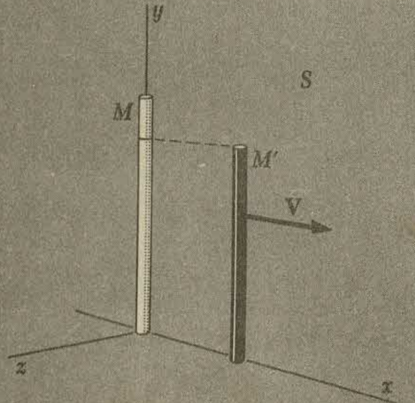
This is the relativistic theory of aberration.



Suppose we have two identical rods M' and M at rest in S.



Suppose M' appears shorter to observer in S when it moves relative to S.



Then we could arrange it so that the end of M' leaves a scratch on M as it passes by. The scratch is a physical result of an experiment, and . . .

Suppose that the particle is a photon, and $v_x' = c$ in S'. From (20a) we see that

$$v_x = \frac{c + V}{1 + cV/c^2} = c. \qquad (20e)$$

The velocity of the photon is also c in the frame S. The Lorentz transformation was designed to produce this result, and it is a reassuring check that we obtain c in both reference frames.

If $v_y = c$ and $v_x = 0$, then

$$v_x' = -V \qquad \text{and} \qquad v_y' = c\left(1 - \frac{V^2}{c^2}\right)^{1/2},$$

so that

$$\frac{v_x'}{v_y'} = -\frac{V}{c(1 - V^2/c^2)^{1/2}}, \qquad (20f)$$

as in Eq. (18).

Example. Velocity addition. Suppose that two particles are traveling opposite to each other with velocity $v_x' = \pm 0.9c$ as observed in the S' system. What is the velocity of one particle with respect to the other? To solve this problem, let S be the reference frame in which the $-0.9c$ particle is at rest. Then the velocity of S' relative to S is $V = 0.9c$ so that the particle which in S' has velocity $v_x' = +0.9c$ has a velocity in S

$$v_x = \frac{v_x' + V}{1 + v_x'V/c^2} \cong \frac{1.8c}{1 + (0.9)^2} = \frac{1.80}{1.81}c = 0.994c. \qquad (21)$$

Notice that the relative velocity of the two particles is less than c.

Example. Velocity addition. If a photon is traveling at velocity $+c$ in S', and S' is traveling relative to S at velocity $+c$, the photon as viewed from S is traveling only at velocity $+c$, and not at $+2c$. This result is contained in (21). The fact of an ultimate speed is a consequence of the structure of the velocity-addition equations which we have derived from the Lorentz transformation. Note further that there is *no* frame in which a photon (light quantum) is at rest.

D. Sadeh has carried out [*Phys. Rev. Letters* **10**, 271 (1963)] a beautiful experiment which shows that the velocity of γ rays is constant (± 10 percent), independent of the velocity of the source, for a source velocity close to $\frac{1}{2}c$ compared with a source at rest. We quote from his paper:

"In our experiments we used the annihilation in flight of positrons. In the annihilation the center-of-mass system of the positron and electron moves with a velocity close to $\frac{1}{2}c$, and two gamma rays are emitted. In the case of annihilation at rest, the two gamma rays are emitted at an angle of 180° and their velocity is c. In the case of annihilation in flight, the angle is smaller than 180° and depends on the energy of the positron. If the velocity of the gamma ray adds on to the velocity of the center of mass according to classical vector addition, and not according to the Lorentz transformation, then the gamma ray traveling with a component of motion in the direction of the positron flight will have a velocity greater than c, and that having a component in the opposite direction will have a velocity smaller than c. If it is found that the two gamma rays reach the counters at the same time for equal distances between the counters and the point of annihilation, this would prove that even for a moving source the two gamma rays travel with the same velocity."

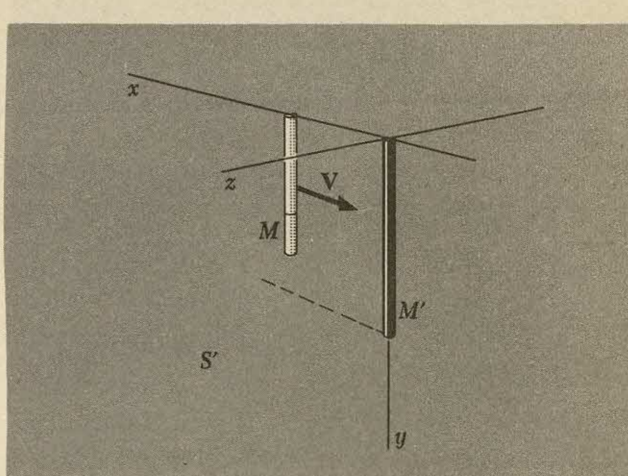
Measurement of Length Perpendicular to Relative Velocity. According to the Lorentz transformation:

$$y' = y; \qquad z' = z. \tag{22}$$

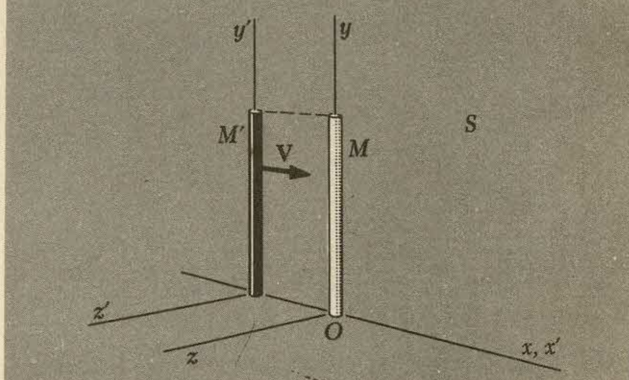
These relations are equivalent to the statement that the measurement of the length of a meter stick is independent of its velocity *if* the meter stick moves perpendicular to its length.

How would we verify this statement experimentally? We can take a meter stick and move it at uniform velocity past another meter stick which is at rest. There is no problem in making the origins of both meter sticks cross exactly. Then the 1-meter mark of each will also cross exactly, or, if the motion changes the length, we can arrange for the 1-meter mark of the shorter stick to make a scratch on the longer stick. This provides a definite physical record of the length.

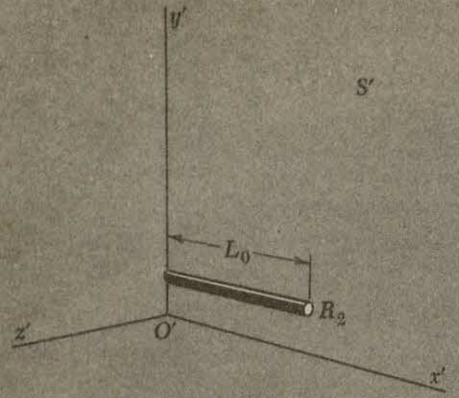
Let S be the rest frame of one meter stick and S′ the rest frame of the other. Suppose the motion does change the apparent length. Then if the laws of physics are to remain the same for an observer on S as for an observer on S′, it is necessary that the stick which appeared the shorter to an observer on S should appear the longer to an observer on S′. But this reversal of the roles is in-



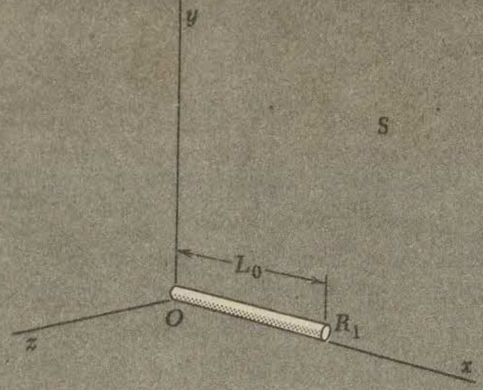
it must be observed in another frame; i.e., upside down in the rest frame of M'. But now M must appear *shorter* than M', since M is moving and M' is at rest.



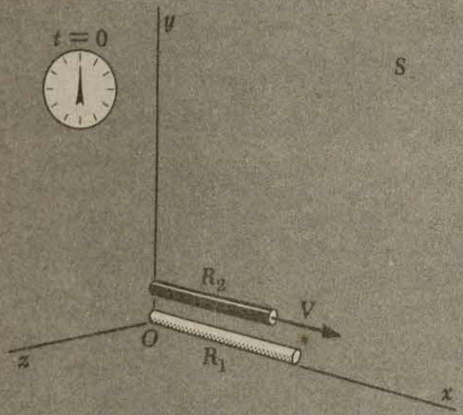
Thus we have a contradiction, which is resolved only if M' and M have the same length even when one of them is moving. Thus $y' = y$. By a similar argument $z' = z$.



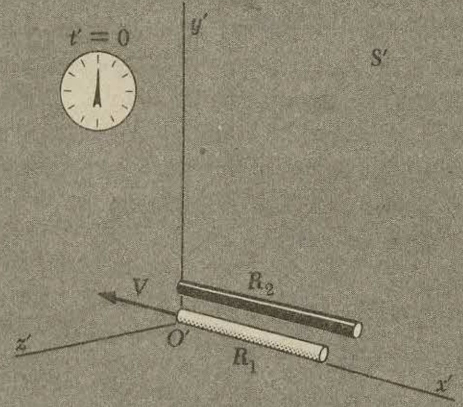
Consider a rigid rod R_2 of length L_0, as measured in its rest frame S′.



A similar rigid rod R_1 of length L_0 in its rest frame S.



The Lorentz transformation tells us that R_2 will be measured to have length $L = L_0\sqrt{1 - V^2/c^2}$ in S. This is the famous Lorentz contraction of moving objects.



In S′, R_2 is at rest, but R_1 has speed V. It will be contracted in S′.

compatible with our physical record that one meter stick is shorter than the other. Therefore, the lengths must be equal when viewed from S and S′. This discussion merely confirms (22).

Length Contraction. Consider a rod lying along the x axis and at rest in the reference frame S. Because the rod is at rest, the position coordinates of its ends, x_1 and x_2, are independent of the time t in S. Thus

$$L_0 = x_2 - x_1 \tag{23}$$

is the rest length of the rod.

Now look at the rod from the reference frame S′ which moves with velocity $V\hat{\mathbf{x}}$ with respect to the rod at rest in S. We determine the length of the rod as viewed from S′ by determining at a given time t' the positions x_1' and x_2' which coincide with the ends of the rod. The distance between positions x_1' and x_2' in S′ which coincide *simultaneously* (in S′) with the end points of the rod is the natural definition of the length L in the moving frame S′:

$$x_2'(t') - x_1'(t') = L. \qquad (24)$$

From the Lorentz transformation (15) we have

$$x_2 = x_2'(t')\gamma + ct'\,\beta\gamma; \qquad x_1 = x_1'(t')\gamma + ct'\,\beta\gamma, \qquad (25)$$

or

$$x_2 - x_1 = L_0 = [x_2'(t') - x_1'(t')]\gamma = L\gamma. \qquad (26)$$

Thus

$$\boxed{L = \frac{L_0}{\gamma} = L_0(1 - \beta^2)^{1/2},} \qquad (27)$$

using our definition $\gamma = (1 - \beta^2)^{-1/2}$. This is the Lorentz-Fitzgerald contraction of a rod moving parallel to its length.

For a discussion of the figures of rapidly moving objects as photographed with a camera, see the excellent review by Weisskopf.[†] It has been shown, for example, by calculation of trajectories that a moving sphere will photograph as a sphere and not as an ellipsoid.

We have seen that there is a difference between a meter stick lying along the y axis and a meter stick lying along the x axis. The result (22) is not the same as (27). With the meter stick along the y axis, we did not have to worry about questions of simultaneity in the comparison of a moving and a stationary meter stick. With the meter stick along the x axis, the matter of simultaneity is all important.

This is illustrated by a different example. We can easily synchronize a series of clocks in S, the frame in which the meter stick is at rest. Let the clocks at $x = 0$ and $x = L_0$ (at each end of the meter stick) each emit at $t = 0$ a directional flash of light in the y direction. These two flashes are received in S′ by two of a series of counters spaced along the x' axis. How far apart are the two counters which were triggered? From (14) we have for the location of the two counters

---

[†] V. F. Weisskopf, *Physics Today* 13, 24–27 (Sept. 1960).

$$x_1' = 0 \cdot \gamma - c \cdot 0 \cdot \beta\gamma = 0; \tag{28}$$

$$x_2' = L_0\gamma - c \cdot 0 \cdot \beta\gamma = L_0\gamma, \tag{29}$$

so that their distance apart is

$$x_2' - x_1' = L_0\gamma = \frac{L_0}{(1 - \beta^2)^{1/2}}. \tag{30}$$

This does not agree with (27)! We have done a *different* experiment and obtained a different result. Our earlier experiment was based on the natural definition of length in S′, using the requirement of simultaneity in S′. The earlier experiment involved comparing $\Delta x'$ with Δx when $\Delta t' = 0$, whereas the second experiment involved comparing $\Delta x'$ with Δx when $\Delta t = 0$.

We have learned indirectly from the result (30) of the second experiment that two events simultaneous in S are not, in general, simultaneous in S′. Thus from (14) we see that two events *simultaneous* ($\Delta t = 0$) in S, which are separated by Δx in space, will be separated in S′ in both space and time:

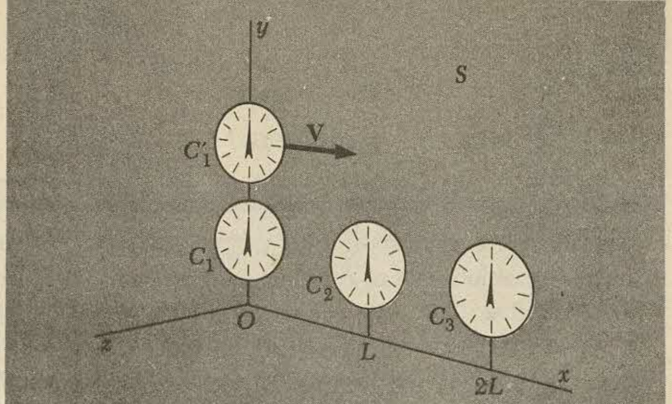
$$\Delta x' = \gamma\,\Delta x; \qquad c\,\Delta t' = -\beta\gamma\,\Delta x.$$

Time Dilation of Moving Clocks. Used in the ordinary sense, the word *dilate* means "enlarge beyond normal size"; in connection with a clock, it means to lengthen an interval of time. We now consider a clock which is at rest in reference frame S. The result of the measurement of a time interval in the frame in which the clock is *at rest* is always denoted by τ and is called the *proper time interval.* Suppose that the clock is situated at the origin $x = 0$ of S. By the Lorentz transformation (14) evaluated at constant x, we have

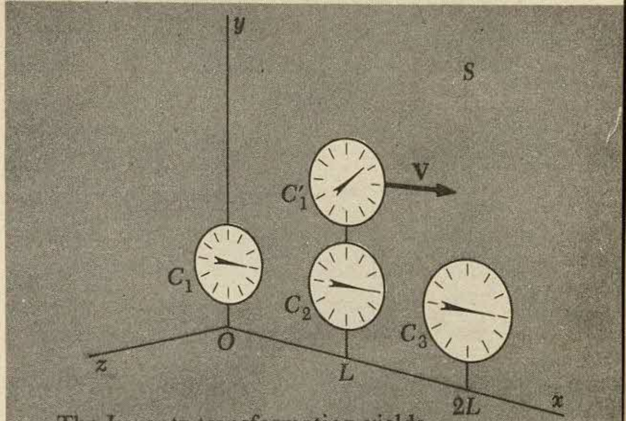
$$\boxed{t' = \tau\gamma = \frac{\tau}{(1 - \beta^2)^{1/2}},} \tag{31}$$

for the time interval measured by a clock in the reference frame S′ moving with velocity $V\hat{\mathbf{x}}$ with respect to the frame S of the original clock. The time interval measured in the moving frame S′ is *longer* than the time interval in the frame S.

This effect is called time dilation. Moving clocks appear to advance more slowly than clocks at rest. (This is not easy to



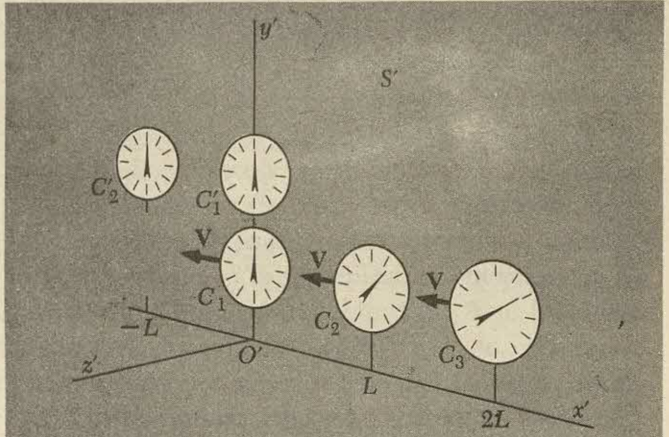
Clocks C_1, C_2, C_3 are at rest in S, spaced at equal intervals L along the x axis, and all synchronized. Clock C_1' has velocity V with respect to S. Suppose $t' = 0$ when $t = 0$, as shown.



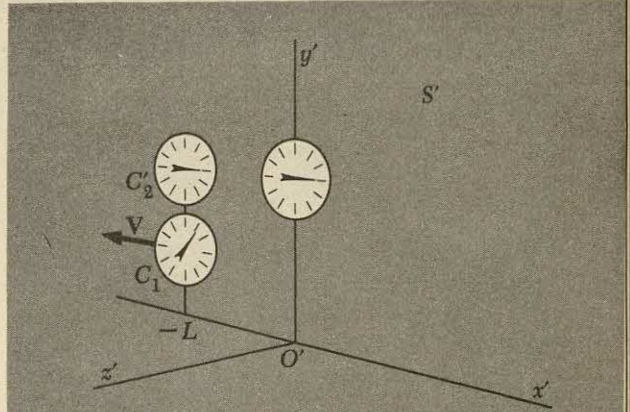
The Lorentz transformation yields

$$t' = (t - xV/c^2)\,\gamma = t\sqrt{1 - V^2/c^2}$$

since $x = L = Vt$. To the observer in S, the moving clock C_1' runs slow.



In S′ clocks C_1', C_2', etc., are at rest, separated by distance L, and synchronized. To the observer in S′, clocks C_1, C_2, C_3 are *not synchronized!* What do they read?

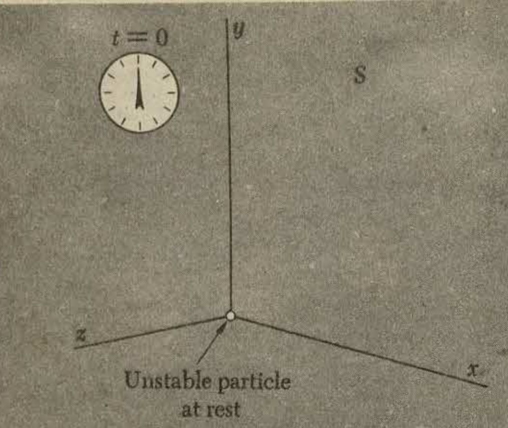


To the observer in S′, it is the *moving clock* C_1 which runs slow! Where are clocks C_2, C_3 and what do they read at this instant?

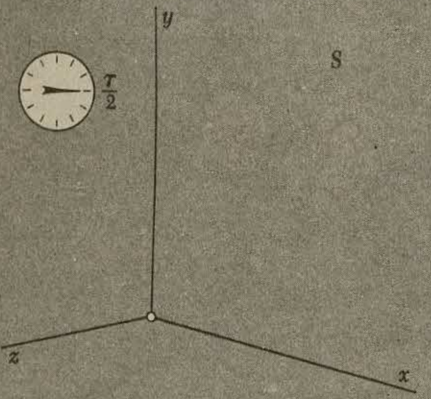
understand in an intuitive way; it may take you a week or a year to feel happy about time dilation. The root of the apparent paradox is the invariance of c.) The effect must hold for any type of clock. In particular, if τ is the decay half-life of mesons or of radioactive matter as measured in the frame S in which the particles are at rest, then

$$t' = \frac{\tau}{(1 - \beta^2)^{1/2}} \tag{32}$$

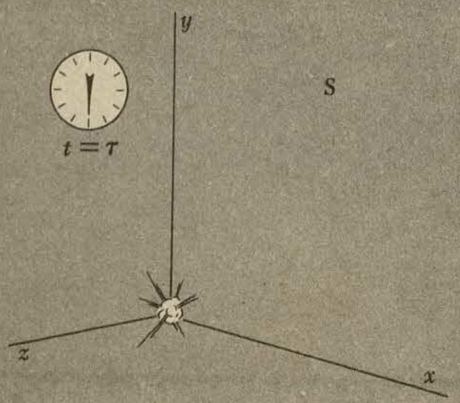
is the decay half-life observed in a frame S′ in which the particles are moving with velocity β.



Another example of time dilation: An unstable particle is at rest in S. We begin to observe it at $t = 0$.



Time elapses . . .



and the particle decays at time $t = \tau$. . . .

Example. Lifetime of π^+ mesons. It is known that a π^+ meson decays into a μ^+ meson and a neutrino. The π^+ meson in a frame in which it is at rest has a mean life before decaying of about 2.5×10^{-8} sec. (Mean lifetimes are discussed in Chap. 15.) If a beam of π^+ mesons is produced with a velocity $\beta \cong 0.9$, what is the lifetime of the beam as viewed from the laboratory reference frame? A π^+ meson is a positively charged unstable particle with mass about $273m$, where m is the mass of the electron. The μ^+ meson has a mass of about $215m$; the neutrino has zero rest mass.

The proper lifetime τ of the π^+ meson is 2.5×10^{-8} sec. If $\beta \cong 0.90$, then $\beta^2 \cong 0.81$ and the expected lifetime in the laboratory frame will be, from (31),

$$t' \cong \frac{2.5 \times 10^{-8}}{(1 - 0.81)^{1/2}} \cong 5.7 \times 10^{-8} \text{ sec.}$$

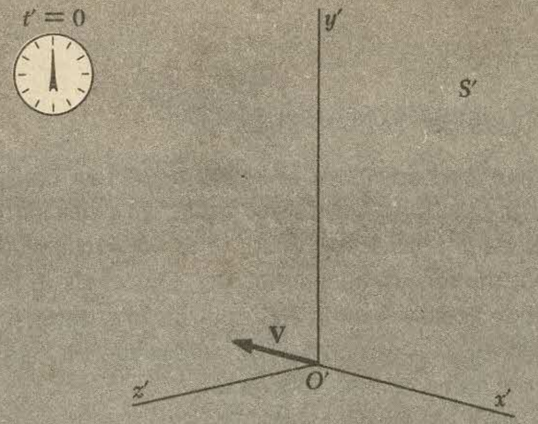
Thus on the average, before decaying, the particle will travel over twice as far as we would expect nonrelativistically from the product of the velocity times the proper lifetime.

Experiments on the lifetime of π^+ mesons (positive pions) are reported by R. P. Durbin, H. H. Loar, and W. W. Havens, Jr., *Phys. Rev.* 88, 179 (1952). The results are in good agreement with the predicted time dilation for the appropriate velocity. Beams of π^+ mesons have been produced with

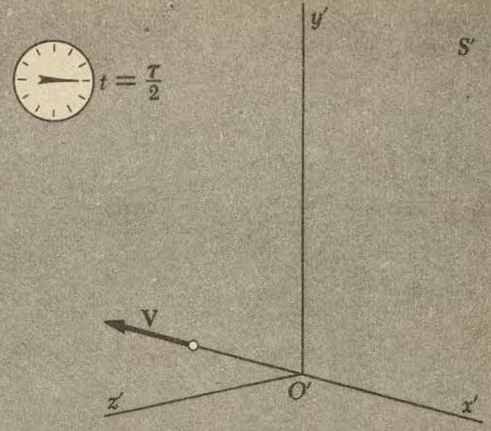
$$\beta = 1 - (5 \times 10^{-5});$$

their mean life in the beam is 2.5×10^{-6} sec, or 100 times the proper lifetime of π^+ mesons at rest.

Consider a beam of π^+ mesons traveling with a velocity nearly equal to c. If the relativistic time-dilation effect did not exist, they would traverse a mean distance equal to $(2.5 \times 10^{-8}$ sec$)$ $(3 \times 10^{10}$ cm/sec$) \approx 700$ cm before decaying. They actually travel much farther than this, because of time dilation. The hydrogen bubble chamber at the Lawrence Radiation Laboratory is about 100 meters from the pion source in the Bevatron. The distance the pions travel before decay is of the order of $(2.5 \times 10^{-6})(3 \times 10^{10}) \approx 10^5$ cm, or about 100 times the distance they would travel before decay without the time-dilation effect. The design of apparatus for high-energy experiments in particle physics takes advantage of the long decay distance due to relativity. It has been said that almost every high-energy



The same phenomenon observed from S'. Now the particle has speed V. We begin to observe it at $t' = 0 = t$.



Time elapses . . . .

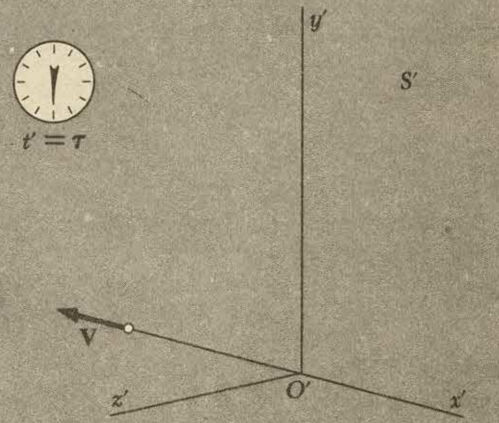
physicist tests special relativity every day. He uses the Lorentz transformation with the same confidence that physicists in the nineteenth century used Newton's laws.

We give a simple discussion which illustrates how the time dilation is forced upon us by the constancy of the speed of light. Let us construct a standard clock in the reference frame S. The clock can be used to measure the time τ needed for a light pulse to travel a fixed distance L from a source at rest to a mirror at rest, and back again. The light path is along the y axis. Thus

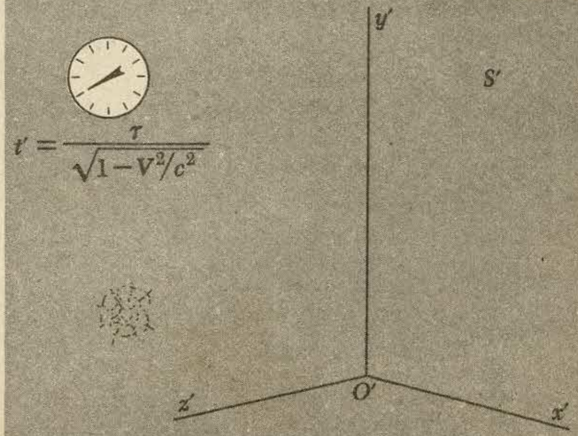
$$\tau = \frac{2L}{c}. \tag{33}$$

This time can be read on a dial or it can be printed out on a piece of paper. Observers in any frame can look at the printed record of the flight time of the pulse and they will all agree that a clock in the rest frame S recorded the time τ. But what do their own clocks, not in S, record? We consider the situation with L in the y direction.

An observer in a frame S′ (moving uniformly in the x direction with respect to S) can also time the light-reflection experiment while it is carried out in S. The observer in S′ will do this by using a set of synchronized clocks at rest in S′. We start two clocks at rest in S′ at the same time (synchronized) by flashing a light source located midway between them; each starts from zero at the instant when the flash reaches it. The procedure may be extended to other clocks. We can also synchronize any number of clocks in one reference frame by synchronizing them when they are close together in space and then separating them slowly until they take up the desired positions.



But at $t' = \tau$ the particle has not yet decayed!



The particle decays at $t' = \tau(1 - V^2/c^2)^{-1/2}$ according to an observer in S'.

We can read any clock in S' and be certain that all other clocks at rest in S' will read the same time. In particular we read whatever clock in S' is closest in space to the single clock in S which is used for the reflection experiment. One clock in S' will be closest and will be read when the light pulse starts out in S; another clock in S' will be closest and will be read when the light pulse returns and is recorded by the clock in S.

The path traversed by the light in S is $2L$. But the path as viewed from S' is longer, because the apparatus in S has moved relative to S' by $V \cdot \frac{1}{2}t'$ along the x axis during the outbound passage of the light pulse from the source to the mirror. The apparatus moves by another $V \cdot \frac{1}{2}t'$ during the inbound passage. Here t' is the time as observed in S'. The distance in S' traveled by the pulse is

$$2[L^2 + (\tfrac{1}{2}Vt')^2]^{1/2};$$

because the pulse travels always with the speed c, this distance must equal ct'. Thus

$$(ct')^2 = 4L^2 + (Vt')^2, \qquad (34)$$

or

$$t' = \frac{2L}{(c^2 - V^2)^{1/2}} = \frac{2L}{c} \cdot \frac{1}{(1 - \beta^2)^{1/2}}, \qquad (35)$$

or

$$t' = \frac{\tau}{(1 - \beta^2)^{1/2}}, \qquad (36)$$

exactly the same as (31). Thus the clock in S will seem to the timekeepers in S' to run slowly, because the S clock has printed out a time τ less than the time t'.

We see that the time-dilation effect does not involve mysterious processes in the interior of atoms; the effect arises in the measurement process. The clock at rest in S reads the proper time τ when viewed by an observer at rest in S; an identical clock at rest in S' will also read τ when viewed by an observer at rest in S'. But when we view from S' a time interval which is τ in S, we see a longer time t' because of the longer light path. Any kind of clock will behave in the same way.

We repeat that there is nothing mysterious about the clocks. If there is anything mysterious about special relativity, it is the constancy of the speed of light. Granted that, everything else follows directly and fairly simply. Every new situation must be analyzed carefully, however. The field is rich in apparent paradoxes. Perhaps the most famous of these is the twin paradox. At the end of the chapter references are given to clear statements and resolutions of the twin paradox.

Example. Longitudinal Doppler effect. Consider two pulses of light sent out at $t = 0$ and $t = \tau$ by a transmitter at rest at $x = 0$ in reference frame S. Reference frame S′ moves with velocity $V\hat{\mathbf{x}}$ with respect to S. The initial pulse is received at $x' = 0$ in S′ at time $t' = 0$. The point in S′ which coincides with $x = 0$ at $t = \tau$ is given by the Lorentz transformation (14):

$$x' = \frac{x - Vt}{(1 - \beta^2)^{1/2}} = \frac{-V\tau}{(1 - \beta^2)^{1/2}}, \tag{37}$$

taking $x = 0$. The corresponding time in S′ is

$$t' = \frac{t - Vx/c^2}{(1 - \beta^2)^{1/2}} = \frac{\tau}{(1 - \beta^2)^{1/2}}. \tag{38}$$

The time needed for the second pulse to travel in S′ from $-V\tau/(1 - \beta^2)^{1/2}$ to the origin is

$$\Delta t' = \frac{\tau V/c}{(1 - \beta^2)^{1/2}}, \tag{39}$$

so that the total time in S′ between the reception at $x' = 0$ of the two pulses is

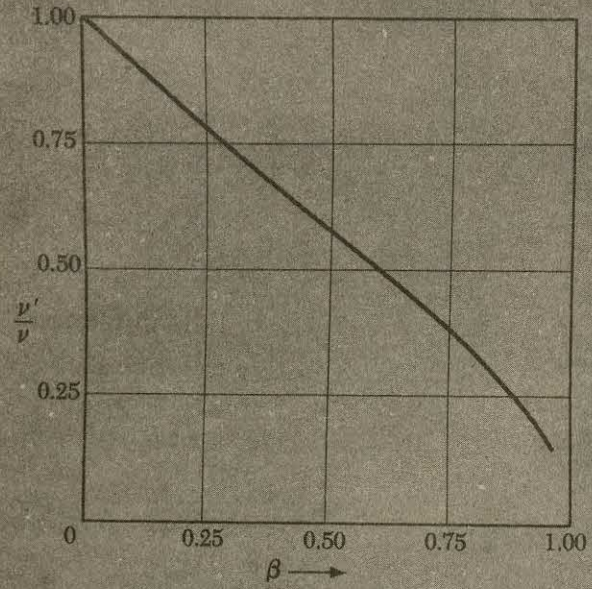
$$t' + \Delta t' = \tau \frac{1 + V/c}{(1 - \beta^2)^{1/2}} = \tau \sqrt{\frac{1 + \beta}{1 - \beta}}. \tag{40}$$

The time between the two signals can equally well be interpreted as the elapsed time between two successive nodes of a light wave. The frequency is the reciprocal of the period of the wave, so that

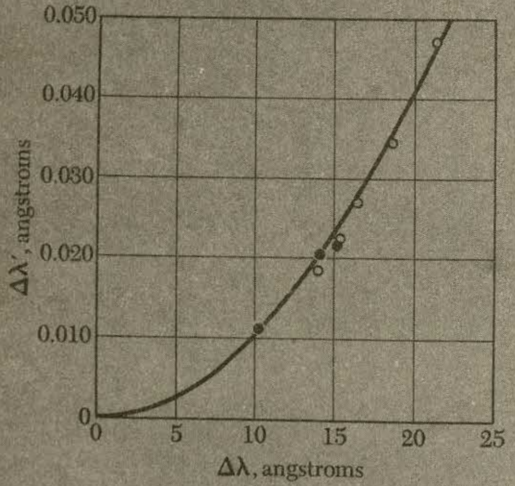
$$\boxed{\nu' = \nu \sqrt{\frac{1 - \beta}{1 + \beta}}.} \tag{41}$$

Here ν' is the frequency as received in S′, and ν is the frequency as transmitted in S. If the receiver is receding from the source, then $\beta = V/c$ is positive and ν' is less than ν. If the receiver is approaching the source, we take β to be negative and ν' is greater than ν. In terms of wavelength, $\lambda = c/\nu$ and $\lambda' = c/\nu'$, so that

$$\lambda' = \lambda \sqrt{\frac{1 + \beta}{1 - \beta}}. \tag{42}$$



Relativistic longitudinal Doppler effect; ν' is the frequency received by an observer moving away from the source at velocity $\beta = V/c$. The source emits the frequency ν.



Computed and observed second-order shifts $\Delta\lambda'$ on vertical axis plotted against first-order (Doppler) shifts $\Delta\lambda$ on horizontal axis, after Ives and Stilwell. Black circles refer to observation on one spectroscopic line and white circles refer to observation on another spectroscopic line.

Equation (41) describes the relativistic longitudinal Doppler effect for light waves in a vacuum. The frequency shift (41) agrees to order β with the nonrelativistic result derived in Chap. 10. The term of order β^2 in the series expansion of (41) has been confirmed experimentally by Ives and Stilwell.

H. E. Ives and G. R. Stilwell [*J. Opt. Soc. Am.* **28,** 215 (1938); **31,** 369 (1941)] have carried out spectroscopic experiments on beams of hydrogen atoms in excited electronic states. The atoms were accelerated as molecular hydrogen ions H_2^+ and H_3^+ in an intense electric field. Atomic hydrogen was formed as a breakup product of the ions. The velocity of the atoms was of the order of $\beta = 0.005$. Ives and Stilwell looked for a shift in the *average* wavelength of a particular spectral line emitted by the hydrogen atoms. The average was taken over the forward and backward directions with respect to the line of flight of the atoms. From (42) we have, using $\beta_{\text{fwd}} = -\beta_{\text{bkwd}}$, the average wavelength

$$\frac{1}{2}(\lambda_{\text{fwd}} + \lambda_{\text{bkwd}}) = \frac{1}{2}\lambda_0\left(\sqrt{\frac{1-\beta}{1+\beta}} + \sqrt{\frac{1+\beta}{1-\beta}}\right)$$
$$= \frac{\lambda_0}{(1-\beta^2)^{1/2}}. \qquad (43)$$

Thus there is a shift of order β^2 in the mean position of the displaced lines, with respect to the wavelength λ_0 emitted from an atom at rest. In their 1941 paper Ives and Stilwell report an observed shift of 0.074 Å in the average wavelength, as compared with the value 0.072 Å calculated from (43) for a value of β deduced from the accelerating potential applied to the original ions. This is an excellent confirmation of the theory of the relativistic Doppler effect.

The *transverse* Doppler effect applies to observations made at right angles to the direction of travel of the light source, which is usually an atom. In the nonrelativistic approximation there is no transverse Doppler effect. A transverse Doppler effect for light waves is predicted by relativity theory; the frequencies must be related as the inverse of the times in (31), so that

$$\nu' = (1 - \beta^2)^{1/2}\nu, \qquad (44)$$

where ν is the frequency in the frame in which the atom is at rest, and ν' is the frequency as observed in a frame moving with velocity V $(=\beta c)$ with respect to the atom.

Accelerated Clocks. The special theory of relativity describes and relates measurements which are independent of the detailed structure of real bodies. It makes no prediction about the dynamical effects of acceleration, such as the stresses induced by acceleration. If such stresses are absent or may be ignored, the theory does give us an unambiguous description of the effect of acceleration on clock rates. The result is as if at each instant an accelerated clock had a different velocity, with a rate to be calculated from (31) with the appropriate instantaneous velocity.

If this prediction is correct, two consequences follow:

1. If the speed is constant but the direction varies, Eq. (31) holds without change. The frame of the clock is noninertial.
2. If the speed is constant except for brief moments of acceleration or deceleration (moments negligibly short in comparison with the total time), then (31) will still describe accurately the relation between the proper time and the stationary laboratory time.

A fast charged particle in a constant magnetic field experiences an acceleration perpendicular to its motion, but the speed never changes. If the particle is unstable, the measured half-life should be exactly the same as if it moved with the same speed in a straight line with no magnetic field present. This forecast is confirmed by experiments on the μ^- meson, which decays with a proper mean life of 2.2×10^{-6} sec into an electron and neutrinos. The same proper lifetime is observed for μ^- mesons which are free or spiraling in a magnetic field or allowed to come to rest. It is believed that the special theory of relativity gives a good description of the circular (accelerated) motion of particles in a magnetic field.

Further reading

M. Born, *Einstein's theory of relativity* (E. P. Dutton, New York, 1924; reprint, Dover, New York, 1962). A patient, full, and clear discussion of the special theory.

H. A. Lorentz, A. Einstein, H. Minkowski, and H. Weyl, *The principle of relativity: A collection of original memoirs,* tr. W. Perrett and G. B. Jeffery (Methuen, London, 1923; reprint, Dover, New York, 1958).

W. H. McCrea, *Relativity physics,* 4th ed. (Wiley, New York, 1954). A readable and concise introduction.

E. P. Ney, *Electromagnetism and relativity* (Harper and Row, New York, 1962). A good brief discussion with emphasis on electromagnetic effects.

W. Pauli, *Theory of relativity,* tr. G. Field (Pergamon, New York, 1958). A translation of an excellent monograph in German (*Relativitätstheorie,* published by B. G. Teubner, Leipzig, 1921). The contents of Part I are not difficult.

R. S. Shankland, "Conversations with Albert Einstein," *Am. J. Phys.* **31,** 47 (1963).

Special relativity theory, selected reprints published for A.A.P.T. (American Institute of Physics, 335 East 45 St., New York 17, New York, 1962). This contains excellent discussions of the famous twin paradox; see especially the papers by Darwin, Crawford, and McMillan.

J. A. Wheeler and E. F. Taylor, *Space-time physics—an introduction* (Freeman, San Francisco, 1965). Highly recommended.

E. Whittaker, *History of the theories of aether and electricity,* 2 vols. (Harper and Row, New York, paperback reprint, 1960).

Film list

"Time Dilation, An Experiment with Mu-mesons" F. Friedman, D. Frisch, and J. Smith (ESI film). Using the radioactive decay of cosmic-ray mu mesons, the dilation of time is shown in an experiment which takes place at Mt. Washington, N. H. (5300 feet) and Cambridge, Mass. (sea level). A detailed report of this experiment appears in *Am. J. Phys.* 31, 342 (1963).

Problems

1. *Lorentz invariant.* Verify from (14) that

$$x^2 - c^2t^2 = x'^2 - c^2t'^2.$$

Note that if we write $x_1 \equiv x$; $x_4 \equiv ict$, then $x^2 - c^2t^2 \equiv x_1^2 + x_4^2$. Here $i = \sqrt{-1}$.

2. *Lorentz transformation.* Given (14), demonstrate (15).

3. *Change of volume.* Show that if L_0^3 is the rest volume of a cube, then

$$L_0^3(1 - \beta^2)^{1/2}$$

is the volume viewed from a reference frame moving with uniform velocity β in a direction parallel to an edge of the cube.

4. *Simultaneity.* Show from the Lorentz transformation that two events simultaneous ($t_1 = t_2$) at different positions ($x_1 \neq x_2$) in reference frame S are not in general simultaneous in reference frame S′.

5. *Addition of velocities.* Show that if in the S′ frame we have $v_y' = c \sin \varphi$ and $v_x' = c \cos \varphi$, then in the S frame

$$v_x^2 + v_y^2 = c^2.$$

The S′ frame moves with velocity $V\hat{\mathbf{x}}$ with respect to the S frame.

6. *π^+ mesons.* (*a*) What is the mean life of a burst of π^+ mesons traveling with $\beta = 0.73$? (The proper mean lifetime τ is 2.5×10^{-8} sec.)

Ans. 3.6×10^{-8} sec.

(*b*) What distance is traveled at $\beta = 0.73$ during one mean life?

Ans. 800 cm.

(*c*) What distance would be traveled without relativistic effects?

Ans. 500 cm.

(*d*) Answer parts (*a*), (*b*), (*c*) again, but for $\beta = 0.99$.

7. *μ mesons.* The proper mean life of the μ meson is approximately 2×10^{-6} sec. Suppose that a large burst of μ mesons produced at some height in the atmosphere travels downward at $v = 0.99c$. The number of collisions in the atmosphere on the way down is small. If 1 percent of those in the original burst survive to reach the earth's surface, estimate the original height. (In the μ meson frame of reference the number of particles which survive to a time t is given by $N(t) = N(0)e^{-t/\tau}$.) *Ans.* 2×10^6 cm.

8. *Two events.* Consider two inertial frames S and S′. Let S′ move with velocity $V\hat{\mathbf{x}}$, with respect to S. At a point x_1' an event takes place at time t_1'. At x_2' another event takes place at time t_2'. The origins coincide at time $t = t' = 0$. Find the corresponding times and distances in S.

9. *π^+ mesons.* A burst of 10^4 π^+ mesons travels in a circular path of radius 20 m at a speed $\beta = 0.99c$. The proper mean life of the π^+ meson is 2.5×10^{-8} sec.

(*a*) How many survive when the burst returns to the point of origin?

(*b*) How many mesons would be left in a burst that had remained at rest at the origin for this same period of time?

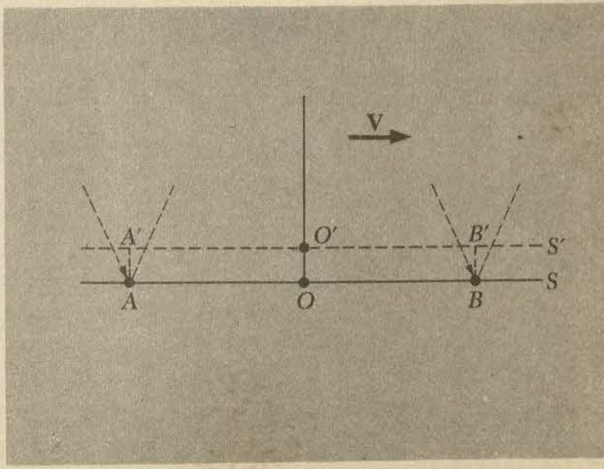
10. *Recessional velocity of galaxy.* We stated in Chap. 10 that red-shift data on distant galaxies gave a velocity of recession proportional to distance, in the nonrelativistic region:

$$V = \alpha r; \qquad \alpha \cong 3 \times 10^{-18}\ \text{sec}^{-1}.$$

Calculate the recession velocity of a galaxy at a distance of 3×10^9 light-years. Is this velocity relativistic? *Ans.* 8.5×10^9 cm/sec.

11. *Galactic velocities.* We observe a galaxy receding in a particular direction at a speed $V = 0.3c$, and another receding in the opposite direction with the same speed. What speed of recession would an oberver in one of these galaxies observe for the other galaxy?

12. *Simultaneity.* Consider the sources of two events to be located at rest at the points A and B, equal distances from the observer O in the frame S. Assume that at the particular instant of time (as determined by observer O in S) at which the two events occur, a second observer O' and his associated reference frame S′, moving with a velocity $V\hat{\mathbf{x}}$ with respect to S, coincide with O and his frame S.



(*a*) Assume $V/c = \frac{1}{3}$. Sketch the positions of the two frames and the points A, A', B, B' when the signal from B' arrives at the observer O'. Has this signal arrived at the observer O? Why?

(*b*) Sketch the positions of S and S′ as both signals arrive at O.

(*c*) Sketch the positions of S and S′ as the signal from A' arrives at O'.

(*d*) Assume that the two events are recorded physically at the points A', B'; for example, on photographic plates. Show under the assumptions of this problem that the distances $A'O'$ and $B'O'$ are equal.

(*e*) Show that the two events are not simultaneous as viewed by O'. The constancy of the velocity of light under all circumstances is implicitly assumed in the definition of simultaneity. To make this dependence clear consider the following. Let the two events at A and B be the simultaneous radiation of pulses of sound as observed by O, an observer at rest with respect to the medium in which the sound is propagated. Let O' be an observer moving with a velocity V one-third that of sound.

(*f*) Use the Galilean transformation to show that the velocity of the sound pulses toward O' from A' and B' are *not* the same as observed by O'.

(*g*) Show that even though the two signals arrive at O' at different times, the fact that the pulses have traveled with different velocities compensates for this fact and that the two events appear simultaneous, even to the observer O'.

13. *Relativistic Doppler shift.* Protons are accelerated through a potential of 20 kv, after which they drift with constant velocity through a region where neutralization to H atoms and associated light emission takes place. The H_β emission ($\lambda = 4861.33$ Å for an atom at rest) is observed in a spectrometer. The optical axis of the spectrometer is parallel to the motion of the ions. The spectrum is Doppler-shifted because of the motion of the ions in the direction of observed emission. The apparatus also contains a mirror which is placed so as to allow superposition of the spectrum of light emitted in the reverse direction. Recall that $1\text{ Å} \equiv 10^{-8}$ cm.

(*a*) What is the velocity of the protons after acceleration?

Ans. 2×10^8 cm/sec.

(*b*) Calculate the first-order Doppler shifts, depending on v/c, appropriate to the forward and backward directions, and indicate the appearance of the relevant part of the spectrum on a diagram.

(*c*) Now consider the second-order, or v^2/c^2, effect which arises from relativistic considerations. Show that the second-order shift is $= \frac{1}{2}\lambda(v^2/c^2)$, and evaluate this numerically for this problem. Notice that it is the same for both $+v$ and $-v$ motions. *Ans.* 0.10 Å.

Mathematical Note: Space-time

Space-time, a mathematical language in which the special theory of relativity finds a particularly simple and elegant expression, was originated by H. Minkowski in 1908 after the special theory of relativity had been proposed. Although it suggests no new results that do not follow also from our previous considerations, it has offered that mathematical form of the theory which is most directly generalizable to the general theory of relativity. Minkowski introduced his contribution as follows: "The views of space and time which I wish to lay before you have sprung from the soil of experimental physics, and therein lies their strength. They are radical. Henceforth space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality."

J. A. Wheeler puts the matter well: "Time is really a length, not an independent concept. To appreciate the falseness of the usual distinction between space and time, consider the inconsistent use of feet to measure the width of a highway and miles to measure its length. Yet it is equally inconsistent to measure intervals in one direction in space-time in seconds, and to measure them in three other directions in centimeters. The conversion factor that changes one metric unit of length (cm) in the 'space' directions to the other metric (sec), still of length, in the 'time' direction is the speed of light—numerically 3×10^{10}. And this factor is just as historical, indeed accidental, in character as the conversion factor 5280 that changes feet to miles! There is no more need to 'explain' 3×10^{10} than there is to 'explain' 5280!"

Space has three dimensions: the position of a particle or any event is specified by three coordinates x, y, and z. We already have a language, the language of vectors, in which we can write statements about relationships among points or lines which are independent of any particular coordinate system. Can we discover an analogous language which tells us how to write physical laws in special relativity in a manner independent of the reference frame? Laws, if expressed without any particular reference frame in mind, will automatically be invariant under the Lorentz transformations which transform us from one reference frame to another.

Physical laws describe the motion of particles from one point to another. Acceleration, collisions, and radioactive decay are examples of phenomena whose description involves both spatial coordinates and time. To represent these laws graphically we imagine a coordinate system with three spatial coordinates x, y, z, and a fourth axis, the time t, perpendicular to the other three. This is hard to visualize, but no harder to use mathematically than three axes. Pictorially we can represent this coordinate system most simply by drawing one space axis x and one time axis t. The three space axes and the one time axis define a four-dimensional continuum called *space-time*. A given point x_0, y_0, z_0, t_0 in space-time is called an *event:* the four coordinates tell us where and when. A typical event may be the collision of a pair of particles. In the prerelativistic world the time of an event t_0 was the same for all observers. In transforming from one inertial system S to another S' the time coordinate t_0 never changed: $t_0 = t_0'$. In the prerelativistic world we had both space *and* time.

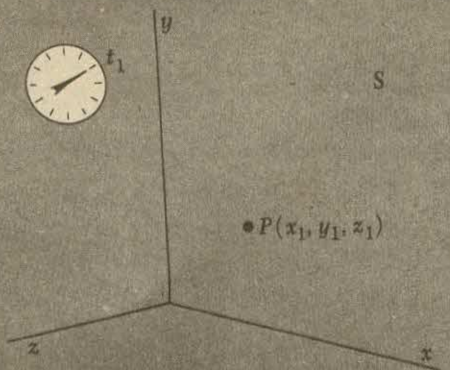
But in relativity the Lorentz transformations mingle the time and spatial coordinates when we transform from one reference frame to another:

$$x' = \frac{x - Vt}{(1 - V^2/c^2)^{1/2}}; \quad y' = y; \quad z' = z; \quad t' = \frac{t - (V/c^2)x}{(1 - V^2/c^2)^{1/2}}. \qquad (45)$$

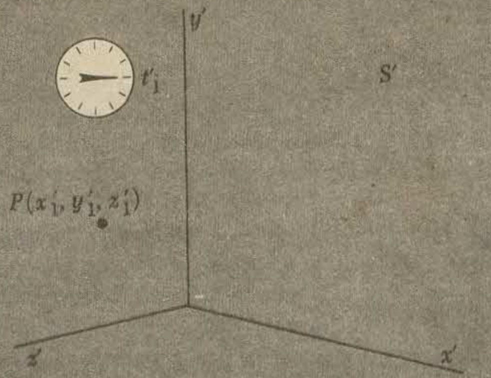
Joining space and time into a single four-dimensional entity, space-time, all of whose dimensions are in a sense equivalent, is an attractive notion. Ordinary rotations mix the spatial coordinate axes only: for a rotation by θ about the z axis we have

$$x' = x\cos\theta + y\sin\theta; \quad y' = -x\sin\theta + y\cos\theta; \quad z' = z; \quad t' = t. \qquad (46)$$

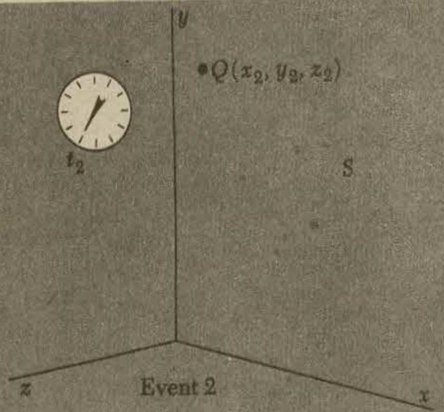
Here x and y are mixed in determining x' and y'. In the Lorentz transformation x and t are mixed. Because no signal propagates faster than the speed



To proceed further we need to define an event: An *event* is defined by the point in space $P(x_1,y_1,z_1)$ and the time t_1 at which it occurs.

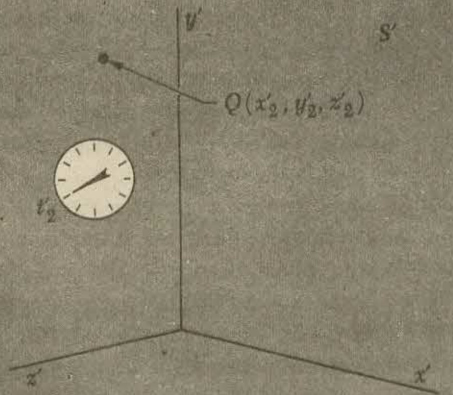


The same event may of course be described by different numbers (x_1', y_1', z_1', t_1') in another inertial frame S′.



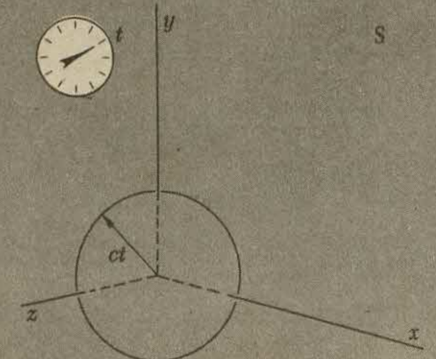
In S the *interval* between events 1 and 2 is defined to be s, where

$$s^2 = c^2(t_2 - t_1)^2 - (x_2 - x_1)^2 - (y_2 - y_1)^2 - (z_2 - z_1)^2.$$



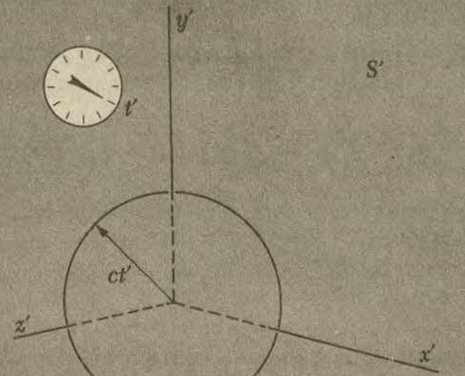
Described in frame S′, the interval between events 1 and 2 is s', where

$$s'^2 = c^2(t_2' - t_1')^2 - (x_2' - x_1')^2 - (y_2' - y_1')^2 - (z_2' - z_1')^2.$$



The interval between emission of a light pulse (at $t = 0$) and its reception at some point (xyz) is $s = 0$:

$$s^2 = c^2t^2 - (x^2 + y^2 + z^2) = c^2t^2 - c^2t^2 = 0.$$



In S′ also, the interval between the events of emission of a light pulse and reception of the pulse is $s' = 0$. Thus in this case $s = s'$.

of light, two events which lie outside each other's light cones can have no effect upon each other. The property of lying outside each other's light cones is independent of the velocity of the reference frame. We know that $x^2 - (ct)^2$ and $(\Delta x)^2 - (c\,\Delta t)^2$ are invariants under a Lorentz transformation. They have the same numerical value in all reference frames moving uniformly with respect to each other:

$$(c\,\Delta t)^2 - (\Delta x)^2 = (c\,\Delta t')^2 - (\Delta x')^2 \equiv (\Delta s)^2. \tag{47}$$

Here Δx and Δt are the differences in space and time coordinates of two events as measured in S; and $\Delta t'$ and $\Delta x'$ are these same differences measured in S'. When two events lie outside each other's light cones in S, then $|c\,\Delta t| < |\Delta x|$ and

$$(c\,\Delta t)^2 - (\Delta x)^2 < 0. \tag{48}$$

Because this function is a Lorentz invariant, the inequality must be true in all reference frames. If two events are on each other's light cones in S, $(\Delta s)^2 \equiv 0$ in both S and S'.

Suppose event 2 lies within the light cone of 1 in S, because $|c\,\Delta t| > |\Delta x|$ for $\Delta t = t_2 - t_1$ and $\Delta x = x_2 - x_1$. Therefore

$$(c\,\Delta t')^2 - (\Delta x')^2 > 0 \tag{49}$$

also in S', and event 2 always lies within the future light cone of 1. Now a signal from 1 can reach the point x_2 at or before the event 2. What happens at 1 can affect what happens at 2, and these events may have some causal relationship. Something that happens at 1 can cause or alter an event 2.

The quantity Δs defined in (47) is called the *interval* between two events. If the events lie outside of each other's light cones, then

$$c^2(\Delta t)^2 < (\Delta x)^2, \tag{50}$$

and the invariant

$$(\Delta s)^2 < 0. \tag{51}$$

In this situation we define the interval as *space-like.* If two events lie within each other's light cones, then

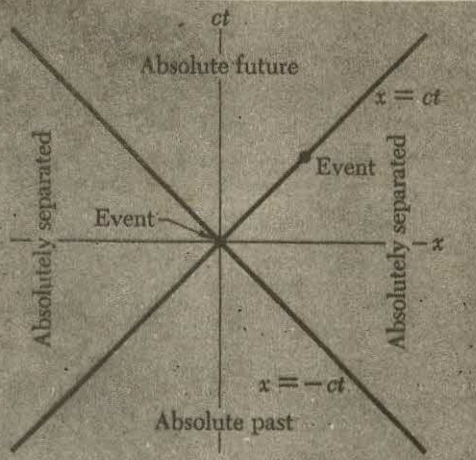
$$(\Delta s)^2 > 0, \tag{52}$$

and we define the interval as *time-like.* For the special situation where the two events can be connected by a light signal, we have

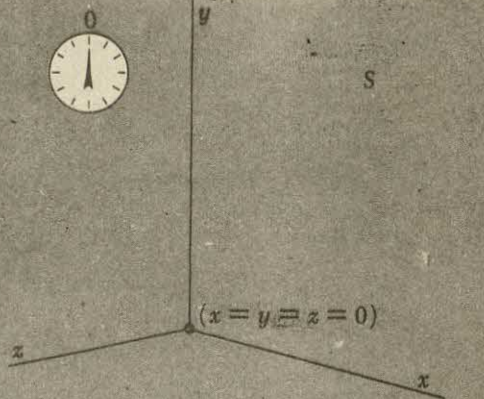
$$(\Delta s)^2 = 0, \tag{53}$$

and the interval is *light-like.*

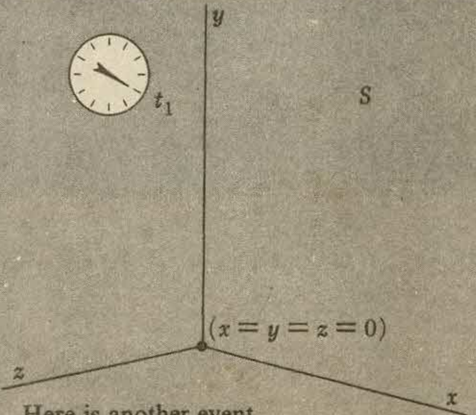
$c\,\Delta t < \|\Delta x\|$	$(\Delta s)^2 < 0$	Space-like	No causal relationship
$c\,\Delta t > \|\Delta x\|$	$(\Delta s)^2 > 0$	Time-like	May have causal relationship
$c\,\Delta t = \|\Delta x\|$	$\Delta s = 0$	Light-like	Can be connected by light signal



A light-cone diagram. The spatial coordinate(s) are represented by x, time by t. All events separated by intervals of zero from the event at $x = 0$, $t = 0$, lie on the lines $x = \pm ct$.

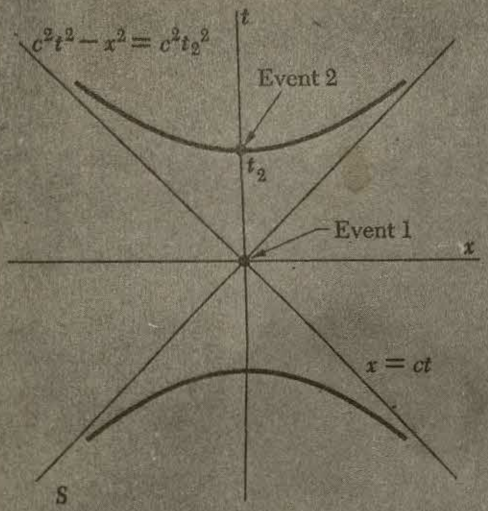


Not all intervals are zero. Here is an event $(x = y = z = t = 0)$.

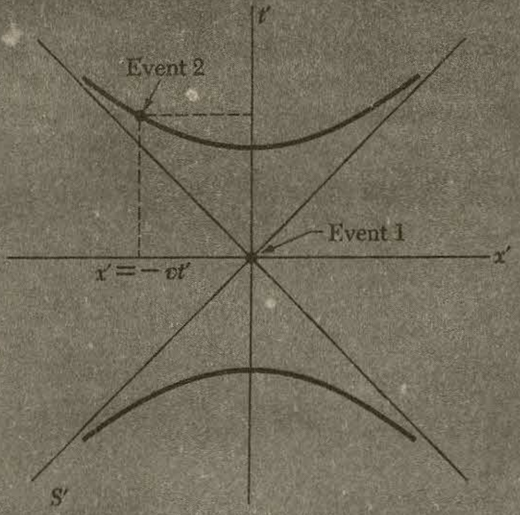


Here is another event $(x = y = z = 0, \quad t = t_1)$.

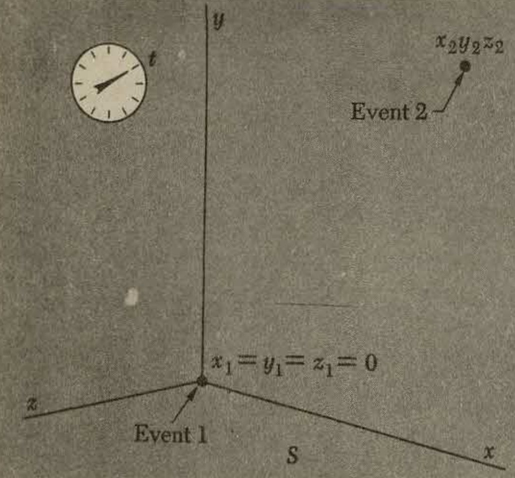
The interval between these two events is s, where $s^2 = c^2 t_1^2 > 0$. Intervals for which $s^2 > 0$ are called *time-like.*



All time-like events separated from event 1 by the same interval s, where $s^2 = c^2t_2^2$, lie on the hyperbola shown.



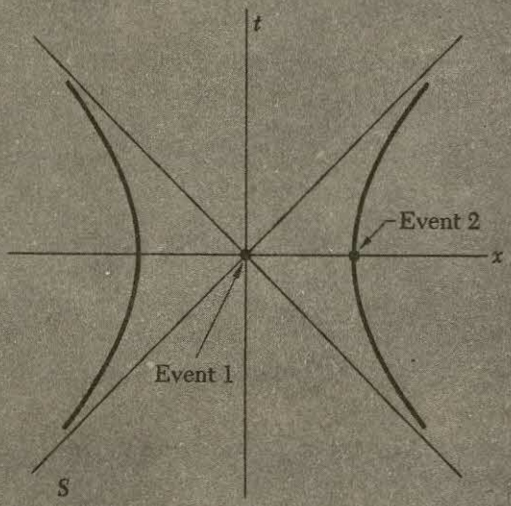
Since the square of the interval s^2 is an invariant in Lorentz transformations, we have the same hyperbola on the light-cone diagram of S′. But event 2 is in a different place. Why?



Another example: Two events with different spatial coordinates but the same time. Here

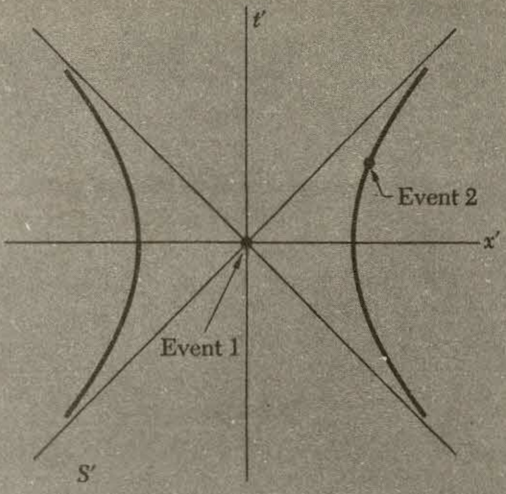
$$s^2 = -(x_2 - x_1)^2 - (y_2 - y_1)^2 - (z_2 - z_1)^2 < 0.$$

Intervals for which $s < 0$ are called space-like.

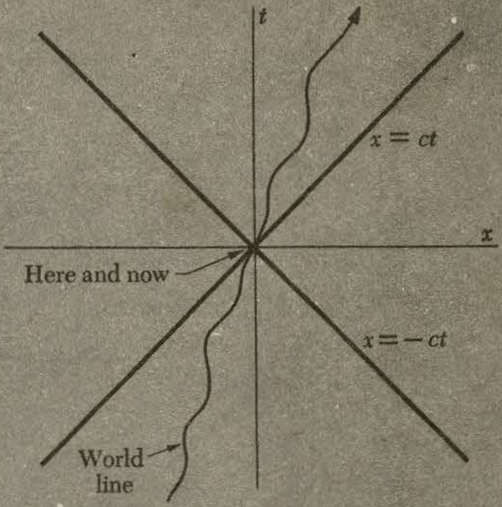


All events separated from event 1 by the same interval as event 2 lie on the hyperbola shown.

Although the concept of simultaneity has no precise meaning in relativity, there is an invariant meaning, the same in all reference systems, to future and past. The *past* consists of all events that could conceivably have had an effect upon us here and now. These events lie in the past light cone. The *future* consists of all events that could conceivably be affected by what we do here and now. These events lie in the future light cone. All events outside the two light cones lie in space-like regions and can neither be



Since the interval s is an invariant, we have the same hyperbola in the S′ diagram. Again, event 2 is no longer in the same position as it was in S.



The "world line" is a succession of events, the later of which are causally connected to the earlier. It must therefore lie within the light cone, as shown.

affected by what we do here and now nor have an effect upon what we do here and now.

Our *world line* is our path in space-time; it is an itinerary giving our position and time. Because we can never exceed the velocity of light, all events which have already occurred in our lives lie in our past cone and all that will occur lie in our future cone.

If instead of one spatial dimension we had considered three, then because $y = y'$ and $z = z'$ in Lorentz transformations with relative velocity V in the x direction, we should again have the invariant squared interval

$$(\Delta s)^2 = (c\,\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2 \tag{54}$$

in S and the same squared interval

$$(\Delta s)^2 = (c\,\Delta t')^2 - (\Delta x')^2 - (\Delta y')^2 - (\Delta z')^2 \tag{55}$$

in some other frame S′.

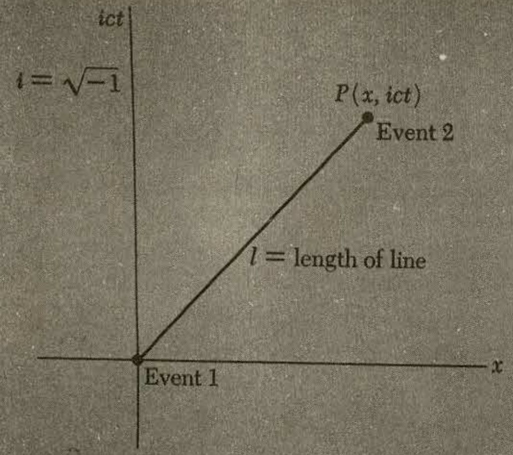
Except for the minus signs, $(\Delta s)^2$ has the same form as the (distance)^2 in Cartesian coordinates between two points of a flat space. And again except for the minus signs, (55) looks like this same (distance)^2 expressed in terms of a different Cartesian coordinate system rotated with respect to the first. This analogy can be exploited in detail in such a way that Lorentz transformations can be treated as a kind of rotation in space-time.

We introduce the coordinates

$$x_1 \equiv x, \qquad x_2 \equiv y, \qquad x_3 \equiv z, \qquad \text{and} \qquad x_4 \equiv ict \tag{56}$$

in analogy to the Cartesian coordinates of a four-dimensional Euclidean space. Here $i = (-1)^{1/2}$. The distance from the origin to a point is

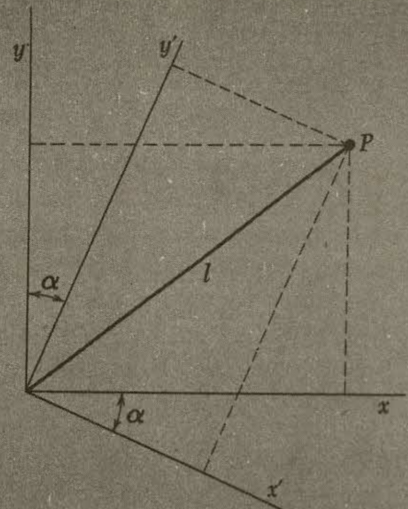
$$x_1^2 + x_2^2 + x_3^2 + x_4^2 \equiv x^2 + y^2 + z^2 - c^2t^2. \tag{57}$$



One way to determine the Lorentz transformation is to take an axis *ict*, and an axis x. Then

$$l^2 = x^2 - c^2t^2.$$

Then $l^2 = -s^2$ is an invariant in the transformation between x',t' and x,t.



When is the distance between two points an invariant in a linear transformation $x,y \rightarrow x',y'$ for ordinary coordinates? *When that transformation is a rotation!*

A rotation about the $x_3 \equiv z$ axis changes the coordinates of a fixed point (a vector) as given in (46). If we express the cosine and the sine of the rotation angle in terms of the ratio of lengths of the sides a, b of a right triangle, then (46) becomes

$$x_1' = \frac{ax_1 + bx_2}{(a^2 + b^2)^{1/2}}; \quad x_2' = \frac{ax_2 - bx_1}{(a^2 + b^2)^{1/2}}. \tag{58}$$

The form of (58) ensures that $x^2 + y^2 = x'^2 + y'^2$:

$$(x_1')^2 + (x_2')^2 = \frac{(a^2 + b^2)(x_1)^2 + (a^2 + b^2)(x_2)^2}{(a^2 + b^2)} = (x_1)^2 + (x_2)^2. \tag{59}$$

But this remains formally true even if $b \rightarrow ib$ and $b^2 \rightarrow -b^2$. Indeed, the Lorentz transformation (9) can be written as

$$x_1' = \frac{ax_1 + ibx_4}{(a^2 - b^2)^{1/2}}; \qquad x_4' = \frac{ax_4 - ibx_1}{(a^2 - b^2)^{1/2}}. \tag{60}$$

Except for the i in front of b this is just like the ordinary rotation (46). In this sense Lorentz transformations are rotations in space-time.

The line from the origin O to the point $P(x,y)$ defines a vector: When the coordinate system is rotated, the coordinates of the end point change in such a way that the vector **OP** remains fixed in space. Similarly, we define the four components $x_1 \equiv x$, $x_2 \equiv y$, $x_3 \equiv z$, $x_4 \equiv ict$ as a vector under Lorentz transformations. These four quantities are generally called a *four-vector.* Similarly, any four numbers which have exactly this same transformation property constitute by definition a four-vector under Lorentz transformations; thus, the four numbers $p_1 \equiv p_x$, $p_2 \equiv p_y$, $p_3 \equiv p_z$, $p_4 \equiv iE/c$ also form a four-vector. Here p refers to momentum and E to energy.

The derivative of a vector with respect to a scalar is a vector. We will see in Chap. 12 that we can write for the μth component of the momentum

$$p_\mu = M\frac{dx_\mu}{d\tau}, \qquad \mu = 1, 2, 3, 4, \tag{61}$$

where τ is the invariant proper time of the particle. The infinitesimal proper time interval $\Delta\tau$ is related to the interval Δs by $(\Delta s)^2 = (c\,\Delta\tau)^2$ so that we also write (61) as

$$p_\mu = Mc\frac{dx_\mu}{ds}, \qquad \mu = 1, 2, 3, 4. \tag{62}$$

The great utility of the idea and notation of a four-vector under Lorentz transformations is that it enables us without thinking to write equations which are independent of any particular unaccelerated reference frame. Such equations satisfy automatically the relativity postulate that general physical laws are the same in all inertial reference frames. With ordinary vectors the equation $\mathbf{a} = \mathbf{b}$ is independent of the coordinate system. In terms of components we have $a_i = b_i$, for $i = 1, 2$, and 3. In a different coordinate system in which the components of $\mathbf{a}$ are a_i' and the components of $\mathbf{b}$ are b_i', it is still true that $a_i' = b_i'$.

We can always form scalar invariants:

$$\mathbf{a}\cdot\mathbf{a} = \sum_{i=1}^{3} a_i a_i = \sum_{i=1}^{3} a_i' a_i', \tag{63}$$

or

$$\mathbf{a}\cdot\mathbf{b} = \sum_{i=1}^{3} a_i b_i = \sum_{i=1}^{3} a_i' b_i'. \tag{64}$$

Similarly, with four-vectors the equation

$$p_\mu^{(1)} + p_\mu^{(2)} = p_\mu^{(3)} + p_\mu^{(4)}, \qquad \mu = 1, 2, 3, 4, \tag{65}$$

holds in all inertial frames for each μ and therefore stands for four equations. It states the conservation of relativistic momentum and mass-energy in the reaction by which the initial particles 1 and 2 become the final particles 3 and 4. The magnitude of the four-vector is a scalar invariant under Lorentz transformations:

$$\sum_{\mu=1}^{4} p_\mu p_\mu = \mathbf{p}\cdot\mathbf{p} - \frac{E^2}{c^2} \equiv -(Mc)^2. \tag{66}$$

This relation is derived in Chap. 12. Similarly,

$$\sum_{\mu=1}^{4} x_\mu x_\mu = x^2 + y^2 + z^2 - (ct)^2 \equiv -s^2 \tag{67}$$

is a scalar, as is

$$\sum_{\mu=1}^{4} x_\mu p_\mu = xp_x + yp_z + zp_z - tE. \tag{68}$$

The sum of two or more vectors is a vector: $P_\mu \equiv p_\mu{}^{(1)} + p_\mu{}^{(2)}$ is a vector, so that, from (66) another scalar is

$$\begin{aligned}\sum_{\mu=1}^{4} p_\mu p_\mu = \sum_{\mu=1}^{4} (p_\mu{}^{(1)} + p_\mu{}^{(2)})^2 &= (\mathbf{p}^{(1)} + \mathbf{p}^{(2)})^2 - \frac{1}{c^2}(E^{(1)} + E^{(2)})^2 \\ &= 2\left[\mathbf{p}^{(1)} \cdot \mathbf{p}^{(2)} - \frac{E^{(1)}E^{(2)}}{c^2}\right] - (M^{(1)}c)^2 - (M^{(2)}c)^2. \end{aligned} \tag{69}$$

In Chap. 12 we arrive at the results of (65) and (69) without any explicit reference to the notion of four-vector in space-time. What we have gained by knowledge of space-time is an extra insight and a simple and elegant way of writing Lorentz-invariant equations. This method is suggestive of generalizations to more abstract and mathematically complicated descriptions, such as arise in relativistic quantum theory and Einstein's general theory of relativity. The ability to write Lorentz-invariant equations without having to prove their invariance enables us to treat more complicated problems than might otherwise be possible.

Lorentz transformations correspond to rotations of the space-time coordinate system. The laws of physics in the special theory of relativity are invariant only under this kind of transformation. Ordinary vectors offer a notation for writing relations without reference to any coordinate system in ordinary three-dimensional space. The Lorentz transformation itself and the simple geometry of space-time have been generalized in a beautiful and remarkable discovery of Einstein. Einstein in the general theory of relativity exploited the possibility of writing physical laws in a form which is independent of *all* transformations in space-time, not just from one unaccelerated reference frame to another. The geometry of space-time is no longer necessarily Euclidean, but can have a curvature.

Historical Note. Simultaneity in special relativity

We give the beginning of the first paper of Einstein on special relativity. The discussion of simultaneity given here is of central importance.

ON THE ELECTRODYNAMICS OF MOVING BODIES†

By A. Einstein

It is known that Maxwell's electrodynamics—as usually understood at the present time—when applied to moving bodies, leads to asymmetries which do not appear to be inherent in the phenomena. Take, for example, the reciprocal electrodynamic action of a magnet and a conductor. The observable phenomenon here depends only on the relative motion of the conductor and the magnet, whereas the customary view draws a sharp distinction between the two cases in which either the one or the other of these bodies is in motion. For if the magnet is in motion and the conductor at rest, there arises in the neighbourhood of the magnet an electric field with a certain definite energy, producing a current at the places where parts of the conductor are situated. But if the magnet is stationary and the conductor in motion, no electric field arises in the neighbourhood of the magnet. In the conductor, however, we find an electromotive force, to which in itself there is no corresponding energy, but which gives rise—assuming equality of relative motion in the two cases discussed—to electric currents of the same path and intensity as those produced by the electric forces in the former case.

Examples of this sort, together with the unsuccessful attempts to discover any motion of the earth relatively to the "light medium," suggest that the phenomena of electrodynamics as well as of mechanics possess no properties corresponding to the idea of absolute rest. They suggest rather that, as has already been shown to the first order of small quantities, the same laws of electrodynamics and optics will be valid for all frames of reference for which the equations of mechanics hold good.* We will raise this conjecture (the purport of which will hereafter be called the "Principle of Relativity") to the status of a postulate, and also introduce another postulate, which is only apparently irreconcilable with the former, namely, that light is always propagated in empty space with a definite velocity *c* which is independent of the state of motion of the emitting body. These two postulates suffice for the attainment of a simple and consistent theory of the electrodynamics of moving bodies based on Maxwell's theory for stationary bodies. The introduction of a "luminiferous ether" will prove to be superfluous inasmuch as the view here to be developed will not require an "absolutely stationary space" provided with special properties, nor assign a velocity-vector to a point of the empty space in which electromagnetic processes take place.

---

† Excerpt from A. Einstein, "Zur Elektrodynamik bewegter Körper," *Ann. Physik* 17, 891–921 (1905), tr. by W. Perrett and G. B. Jeffery, in *The principle of relativity* (Methuen, London, 1923; reprint, Dover, New York). All other footnotes are also excerpted.

* The preceding memoir by Lorentz was not at this time known to the author.

The theory to be developed is based—like all electrodynamics—on the kinematics of the rigid body, since the assertions of any such theory have to do with the relationships between rigid bodies (systems of co-ordinates), clocks, and electromagnetic processes. Insufficient consideration of this circumstance lies at the root of the difficulties which the electrodynamics of moving bodies at present encounters.

I. KINEMATICAL PART

§1. Definition of Simultaneity

Let us take a system of co-ordinates in which the equations of Newtonian mechanics hold good.* In order to render our presentation more precise and to distinguish this system of co-ordinates verbally from others which will be introduced hereafter, we call it the "stationary system."

If a material point is at rest relatively to this system of co-ordinates, its position can be defined relatively thereto by the employment of rigid standards of measurement and the methods of Euclidean geometry, and can be expressed in Cartesian co-ordinates.

If we wish to describe the *motion* of a material point, we give the values of its co-ordinates as functions of the time. Now we must bear carefully in mind that a mathematical description of this kind has no physical meaning unless we are quite clear as to what we understand by "time." We have to take into account that all our judgments in which time plays a part are always judgments of *simultaneous events*. If, for instance, I say, "That train arrives here at 7 o'clock," I mean something like this: "The pointing of the small hand of my watch to 7 and the arrival of the train are simultaneous events."†

It might appear possible to overcome all the difficulties attending the definition of "time" by substituting "the position of the small hand of my watch" for "time." And in fact such a definition is satisfactory when we are concerned with defining a time exclusively for the place where the watch is located; but it is no longer satisfactory when we have to connect in time series of events occurring at different places, or—what comes to the same thing—to evaluate the times of events occurring at places remote from the watch.

---

* i.e. to the first approximation.

† We shall not here discuss the inexactitude which lurks in the concept of simultaneity of two events at approximately the same place, which can only be removed by an abstraction.

We might, of course, content ourselves with time values determined by an observer stationed together with the watch at the origin of the co-ordinates, and co-ordinating the corresponding positions of the hands with light signals, given out by every event to be timed, and reaching him through empty space. But this co-ordination has the disadvantage that it is not independent of the standpoint of the observer with the watch or clock, as we know from experience. We arrive at a much more practical determination along the following line of thought.

If at the point A of space there is a clock, an observer at A can determine the time values of events in the immediate proximity of A by finding the positions of the hands which are simultaneous with these events. If there is at the point B of space another clock in all respects resembling the one at A, it is possible for an observer at B to determine the time values of events in the immediate neighbourhood of B. But it is not possible without further assumption to compare, in respect of time, an event at A with an event at B. We have so far defined only an "A time" and a "B time." We have not defined a common "time" for A and B, for the latter cannot be defined at all unless we establish *by definition* that the "time" required by light to travel from A to B equals the "time" it requires to travel from B to A. Let a ray of light start at the "A time" t_A from A towards B, let it at the "B time" t_B be reflected at B in the direction of A, and arrive again at A at the "A time" t_A'.

In accordance with definition the two clocks synchronize if

$$t_B - t_A = t_A' - t_B.$$

We assume that this definition of synchronism is free from contradictions, and possible for any number of points; and that the following relations are universally valid:—

1. If the clock at B synchronizes with the clock at A, the clock at A synchronizes with the clock at B.
2. If the clock at A synchronizes with the clock at B and also with the clock at C, the clocks at B and C also synchronize with each other.

Thus with the help of certain imaginary physical experiments we have settled what is to be understood by synchronous stationary clocks located at different places, and have evidently obtained a definition of "simultaneous," or "synchronous," and of "time." The "time" of an event is that which is given simultaneously with the event by a stationary clock located at the place of the event, this clock being synchronous, and indeed synchronous for all time determinations, with a specified stationary clock.

In agreement with experience we further assume the quantity

$$\frac{2\mathrm{AB}}{t_{\mathrm{A}}' - t_{\mathrm{A}}} = c,$$

to be a universal constant—the velocity of light in empty space.

It is essential to have time defined by means of stationary clocks in the stationary system, and the time now defined being appropriate to the stationary system we call it "the time of the stationary system."

§2. *On the Relativity of Lengths and Times*

The following reflexions are based on the principle of relativity and on the principle of the constancy of the velocity of light. These two principles we define as follows:—

1. The laws by which the states of physical systems undergo change are not affected, whether these changes of state be referred to the one or the other of two systems of co-ordinates in uniform translatory motion.

2. Any ray of light moves in the "stationary" system of co-ordinates with the determined velocity c, whether the ray be emitted by a stationary or by a moving body. Hence

$$\text{velocity} = \frac{\text{light path}}{\text{time interval}}$$

where time interval is to be taken in the sense of the definition in §1.

Let there be given a stationary rigid rod; and let its length be l as measured by a measuring-rod which is also stationary. We now imagine the axis of the rod lying along the axis of x of the stationary system of co-ordinates, and that a uniform motion of parallel translation with velocity v along the axis of x in the direction of increasing x is then imparted to the rod. We now inquire as to the length of the moving rod, and imagine its length to be ascertained by the following two operations:—

(*a*) The observer moves together with the given measuring-rod and the rod to be measured, and measures the length of the rod directly by superposing the measuring-rod, in just the same way as if all three were at rest.

(*b*) By means of stationary clocks set up in the stationary system and synchronizing in accordance with §1, the observer ascertains at what points of the stationary system the two ends of the rod to be measured are located at a definite time. The distance between these two points, measured by the measuring-rod already employed, which in this case is at rest, is also a length which may be designated "the length of the rod."

In accordance with the principle of relativity the length to be discovered by the operation (*a*)—we will call it "the length of the rod in the moving system"—must be equal to the length l of the stationary rod.

The length to be discovered by the operation (*b*) we will call "the length of the (moving) rod in the stationary system." This we shall determine on the basis of our two principles, and we shall find that it differs from l.

Current kinematics tacitly assumes that the lengths determined by these two operations are precisely equal, or in other words, that a moving rigid body at the epoch t may in geometrical respects be perfectly represented by *the same* body *at rest* in a definite position.

We imagine further that at the two ends A and B of the rod, clocks are placed which synchronize with the clocks of the stationary system, that is to say that their indications correspond at any instant to the "time of the stationary system" at the places where they happen to be. These clocks are therefore "synchronous in the stationary system."

We imagine further that with each clock there is a moving observer, and that these observers apply to both clocks the criterion established in §1 for the synchronization of two clocks. Let a ray of light depart from A at the time* t_A, let it be reflected at B at the time t_B, and reach A again at the time t_A'. Taking into consideration the principle of the constancy of the velocity of light we find that

$$t_B - t_A = \frac{r_{AB}}{c - v} \qquad \text{and} \qquad t_A' - t_B = \frac{r_{AB}}{c + v}$$

where r_{AB} denotes the length of the moving rod—measured in the stationary system. Observers moving with the moving rod would thus find that the two clocks were not synchronous, while observers in the stationary system would declare the clocks to be synchronous.

So we see that we cannot attach any *absolute* signification to the concept of simultaneity, but that two events which, viewed from a system of co-ordinates, are simultaneous, can no longer be looked upon as simultaneous events when envisaged from a system which is in motion relatively to that system. . . .

---

* "Time" here denotes "time of the stationary system" and also "position of hands of the moving clock situated at the place under discussion."

Chapter 12

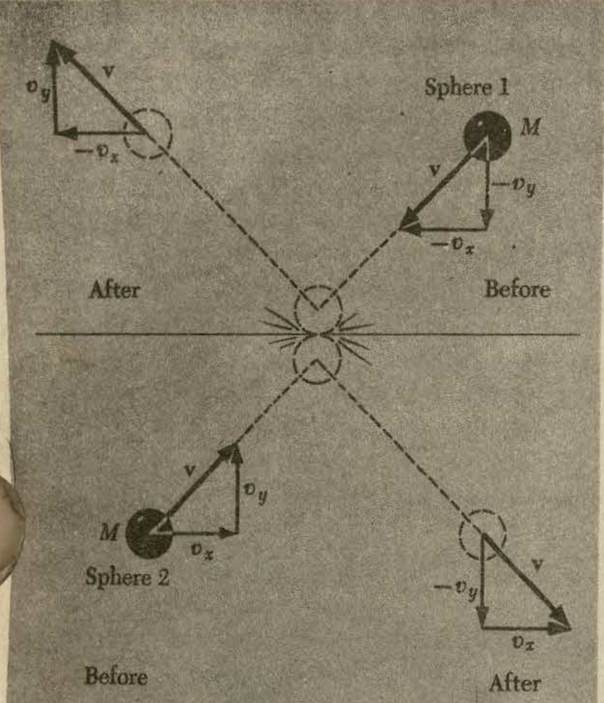
Conservation of momentum 382
Relativistic energy 386
Transformation of momentum and energy 388
Example: Inelastic collision 389
Equivalence of mass and energy 390
Examples: Mass-energy conversions 391
Stellar energy reactions 394
Work and energy 395
Particles with zero rest mass 397
Transformation of rate of change of momentum 400
Constancy of charge 402
Problems 402
Historical Note: Mass-energy relation 404

Relativistic Dynamics: Momentum and Energy

The basic change in our concepts of space and time as expressed in the Lorentz transformation deeply affects all of physics. We must reexamine now the laws of physics as developed and confirmed for low velocities ($v \ll c$) to see if they are compatible with the theory of relativity. We should not be surprised to find that the laws change when carried into new domains. The laws change in such a way that at low velocities we retrieve the Newtonian forms, which we know from experience are accurate in the limit of low velocities.

As in Chap. 3, we accept as possible physical laws only those which are identical in all reference frames in unaccelerated relative motion. But instead of the Galilean transformation to tell us how to transform a physical law from one reference frame to another, we now use the Lorentz transformation. The Lorentz transformation reduces to the Galilean when $v/c \rightarrow 0$. In place of insisting upon the invariance of physical laws under Galilean transformations, we now insist upon their invariance under Lorentz transformations.

Two observers in different reference frames S and S′ deduce physical laws. Each expresses them in terms of lengths, times, velocities, or accelerations, as measured in his own system. The laws must be identical in form in the S frame variables and in the S′ frame variables. Thus when we use the Lorentz transformation to transform from the x, y, z, t of S to the x', y', z', t' of S′, any physical law deduced in S is translated to the language of S′ and should be unchanged in form. The meaning of this will become clear as we examine particular problems.

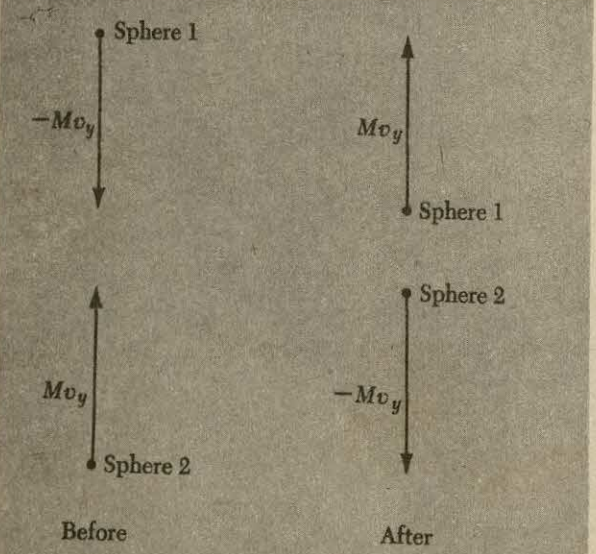


A collision between two spheres of mass M, taking place in the xy plane. The velocities in the x and y directions before and after collision are as shown.

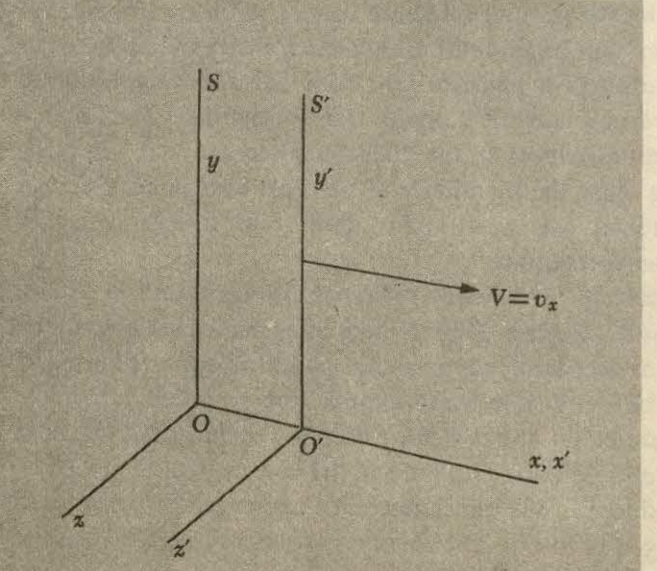
Conservation of Momentum. We want to find a definition of the momentum $\mathbf{p}$ which reduces to $M\mathbf{v}$, where M is the rest mass,[†] for $v/c \ll 1$, and which assures momentum conservation in collisions regardless of the velocities of the particles relative to the reference frame. We will find the appropriate definition by consideration of a particular collision. We first show by an example that the Newtonian (nonrelativistic) momentum $M\mathbf{v}$ is not conserved in collisions involving *relativistic* velocities.

Consider the accompanying figures which describe a collision between particles of *equal mass.* We choose a reference frame S such that the particles approach each other with equal and opposite velocities: the y velocity component of particle 1 is $-v_y$ before the collision and $+v_y$ after the collision. In this reference frame the center of mass is at rest. The total y component of momentum must be zero by symmetry, both before and after the collision. This will be true no matter what definition we use for momentum, provided that it has opposite signs for $\pm v_y$. We therefore encounter no trouble here (whether or not the expression be correct) with the Newtonian definition $\mathbf{p} = M\mathbf{v}$: the change in p_y of particle 1 is $+2Mv_y$, and the change in p_y of particle 2 is $-2Mv_y$, so that the total change in the y component of the Newtonian momentum is zero.

[†] The rest mass is defined as the inertial mass in the nonrelativistic limit $v/c \ll 1$.



The individual nonrelativistic momenta in the y direction are shown. The *total* momentum in the y direction is zero before *and* after the collision.



We have viewed the collision in frame S. What if we view it in frame S', which has the particular velocity $V = v_x$ with respect to S, as shown?

In S′, we find

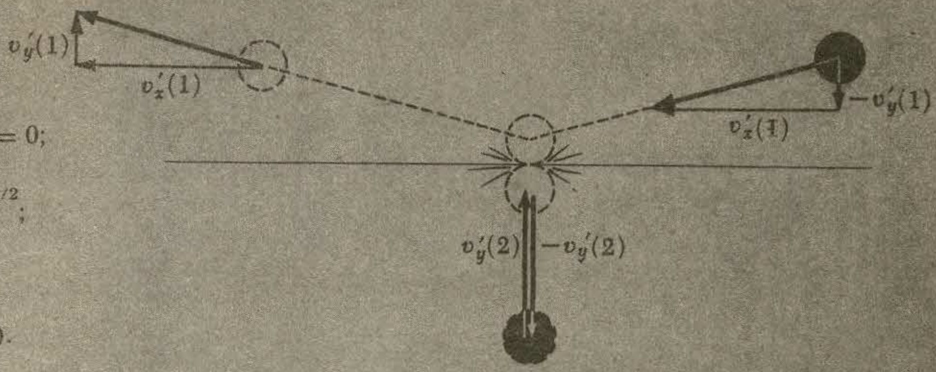
$$v_x'(1) = -\frac{2V}{1 + V^2/c^2}; \qquad v_x'(2) = 0;$$

$$v_y'(1) = \frac{v_y}{1 + V^2/c^2}\left(1 - \frac{V^2}{c^2}\right)^{1/2};$$

and

$$v_y'(2) = \frac{v_y}{(1 - V^2/c^2)^{1/2}} > v_y'(1).$$

[See Eq. (11.24).]



The primed reference frame S′ moves with the particular velocity $\mathbf{V} = v_x\hat{\mathbf{x}}$ with respect to the unprimed reference frame S; here v_x is the x component of the velocity of particle 2 before the collision. We know from the relation for the relativistic addition of velocities in Eq. (11.20*d*) that a particle velocity component seen as v_y by an observer at rest in S will be seen as

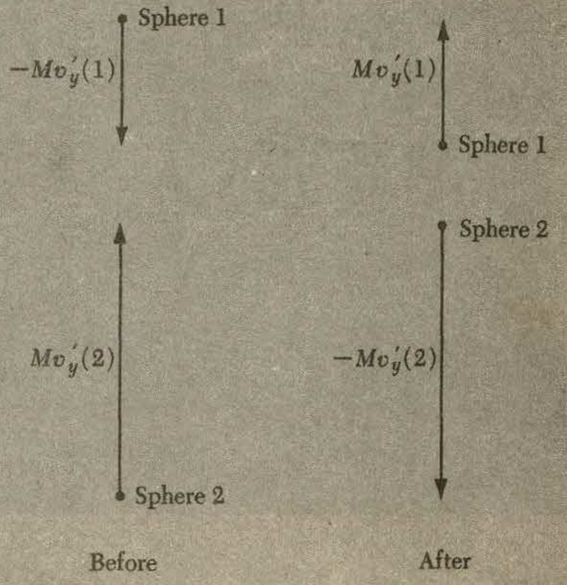
$$v_y' = \frac{v_y}{1 - v_xV/c^2}\left(1 - \frac{V^2}{c^2}\right)^{1/2} \tag{1}$$

by an observer at rest in S′. This result has the consequence that the magnitude of the y components of the velocities of particles 1 and 2 are not equal in S′, even though they were equal in S. For, we see from (1) with $V = v_x$ (of particle 2) that we have for the initial y velocity components as viewed in S′:

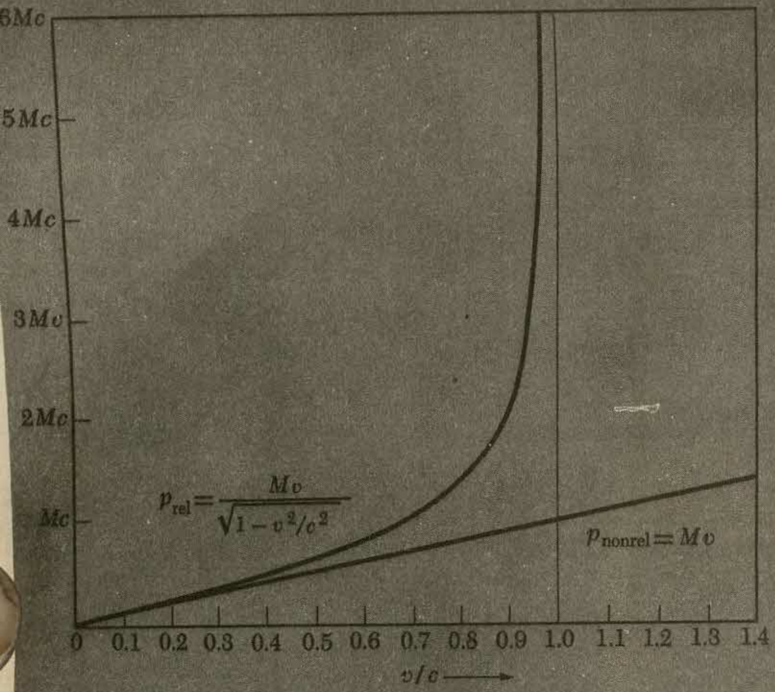
$$-v_y'(1) = -\frac{v_y}{1 + v_x^2/c^2}\left(1 - \frac{v_x^2}{c^2}\right)^{1/2}; \tag{2}$$

$$v_y'(2) = \frac{v_y}{1 - v_x^2/c^2}\left(1 - \frac{v_x^2}{c^2}\right)^{1/2}. \tag{3}$$

These velocity components are now not equal in magnitude. Because the x components of velocity of particles 1 and 2 are not equal in S (they are the negatives of each other), it follows from (1) that the y components of velocity have different magnitudes as seen from any reference frame moving relative to S in the x direction. Thus the momentum change $2Mv_y'(2)$ is not equal in magnitude to the momentum change $2Mv_y'(1)$. We see that a definition in which the momentum is directly proportional to velocity cannot assure momentum conservation in all reference



In the new frame S′ the nonrelativistic momentum is *not* the same in the y' direction before and after collision. There is a net increase in the y component of the nonrelativistic momentum.



So that momentum conservation will hold in *all* frames, we redefine $\mathbf{p}$ as follows: For a particle with velocity $\mathbf{v}$ and rest mass M,

$$\mathbf{p} = \frac{M\mathbf{v}}{\sqrt{1 - v^2/c^2}}.$$

The magnitudes of both the relativistic momentum and the nonrelativistic momentum are plotted in the graph.

systems. Either momentum conservation is incompatible with Lorentz invariance or there exists another definition of momentum such that momentum conservation is valid in all systems with constant relative velocities.

We now look for a definition of momentum which is Lorentz invariant. The definition must be such that the y component of the momentum of a particle is independent of the x component of the velocity of the reference frame in which the collision is observed. If such a definition is found, then conservation of the y component of momentum in one reference frame ensures its conservation in all reference frames. We know that under Lorentz transformations the displacement Δy in the y direction is the same for all reference frames. But the time Δt to go the distance Δy depends upon the reference frame, and thus the velocity component $v_y = \Delta y/\Delta t$ depends on the reference frame. Instead of laboratory clocks to measure Δt, we can refer to a clock carried on the particle. This clock measures the proper time interval $\Delta\tau$ of the particle. *All observers will agree on the value of* $\Delta\tau$. Thus the quantity $\Delta y/\Delta\tau$ is the same in all the reference frames.

We know that Δt and $\Delta\tau$ differ by the time-dilation factor Eq. (11.31), so we have

$$\Delta\tau = \Delta t\left(1 - \frac{v^2}{c^2}\right)^{1/2}, \tag{4}$$

whence

$$\frac{\Delta y}{\Delta\tau} = \frac{\Delta y}{\Delta t}\cdot\frac{\Delta t}{\Delta\tau} = \frac{\Delta y}{\Delta t}\,\frac{1}{(1 - v^2/c^2)^{1/2}}. \tag{5}$$

We see from (5) that the y component of $\mathbf{v}/(1 - v^2/c^2)^{1/2}$ will be the same in all reference frames which differ only in their x component of velocity. If we *define* the relativistic momentum by

$$\boxed{\mathbf{p} \equiv \frac{M\mathbf{v}}{(1 - v^2/c^2)^{1/2}},} \tag{6}$$

then conservation of the y component of momentum is valid in any other inertial reference frame which differs from the rest frame by a constant velocity in the x direction. Note that we may write

$$p = Mc\beta\gamma, \tag{7}$$

from the definitions $\beta = v/c$ and $\gamma = (1 - v^2/c^2)^{-1/2}$ introduced in Chap. 11.

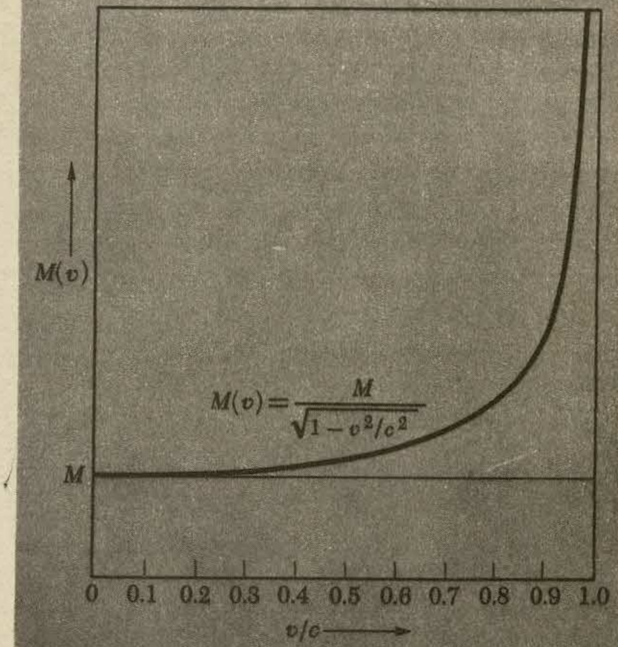
For convenience we have used axes arranged in a symmetrical way so that in this collision there is no change in the x component of velocity of either particle. It follows that p_x as defined by (6) is automatically conserved. Thus with the definition (6) we have complete relativistic momentum conservation in the two-body collision of identical particles. The student may show that even if particle 2 has a different mass from particle 1, the above argument holds, so that we have a relativistic law of momentum conservation. For $v/c \ll 1$ the definition of momentum reduces to the nonrelativistic result $\mathbf{p} = M\mathbf{v}$. It is an experimental fact that the momentum as defined by (6) is conserved in all collision processes.

We may write the relativistic momentum (6) as

$$\mathbf{p} = M(v)\,\mathbf{v}, \tag{8}$$

so that we may interpret

$$M(v) \equiv \frac{M}{(1 - v^2/c^2)^{1/2}} = M\gamma \tag{9}$$



The new definition of momentum leads to this behavior of the mass:

$$M(v) = \frac{M}{\sqrt{1 - v^2/c^2}}.$$

as the relativistic mass of the particle of rest mass M when in motion with the speed v. The rest mass is the mass for $v \to 0$. As $v \to c$, $M(v)/M \to \infty$. The relativistic increase of mass has been verified in various electron deflection experiments; it is also verified implicitly in the operation of every high-energy particle accelerator. The alternate formulation to Eq. (9) which is given below emphasizes directly the relation between relativistic energy and momentum, and it is often simpler to apply.

Relativistic Energy. We first consider the relativistic energy from a formal viewpoint. From (7) the square of the relativistic momentum can be written as

$$p^2 = M^2c^2\,\beta^2\gamma^2. \tag{10}$$

The identity

$$\frac{1}{1 - v^2/c^2} - \frac{v^2/c^2}{1 - v^2/c^2} = 1, \tag{11}$$

or

$$\gamma^2 - \beta^2\gamma^2 = 1, \tag{11a}$$

is a ready-made Lorentz invariant, because 1 is a constant. On multiplying by M^2c^4, we have

$$M^2c^4(\gamma^2 - \beta^2\gamma^2) = M^2c^4, \tag{12}$$

or

$$M^2c^4\gamma^2 - p^2c^2 = M^2c^4, \qquad \text{using (10)}. \tag{12a}$$

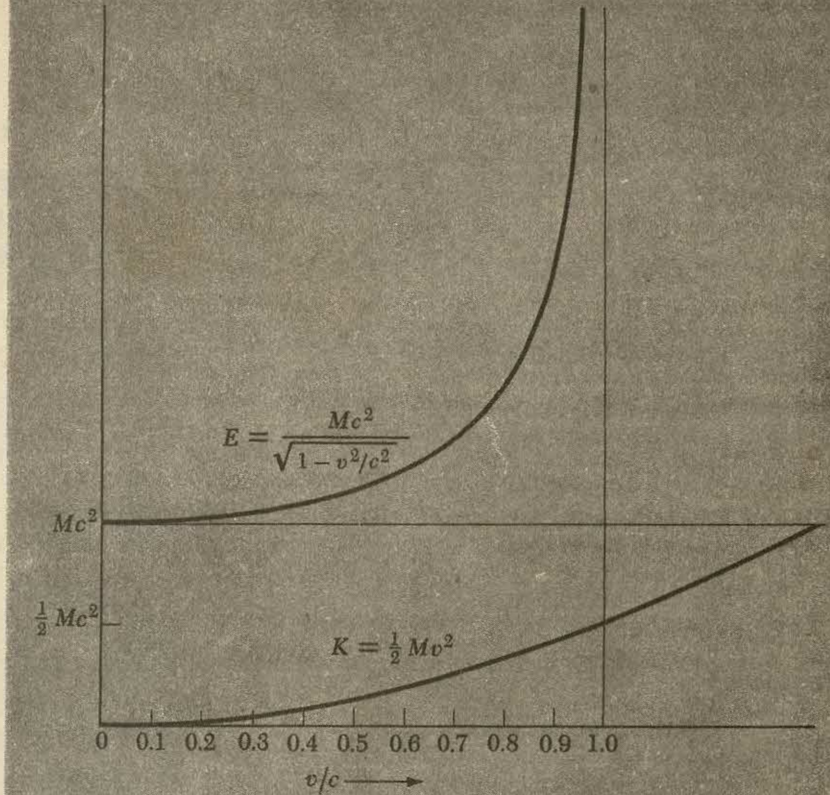
Because the *rest mass* is a constant, we know that M^2c^4 is a constant and therefore a Lorentz invariant as required. But what physical quantity is $M^2c^4\gamma^2$? Its role in (12*a*) strongly suggests that it must be an important physical quantity, for when p^2c^2 is subtracted from it, we have a number (M^2c^4) which is invariant under Lorentz transformations.

Let us consider the quantity

$$Mc^2\gamma = Mc^2\frac{1}{(1 - \beta^2)^{1/2}} \tag{13}$$

For $\beta \ll 1$, this becomes

$$Mc^2\gamma \cong Mc^2(1 + \tfrac{1}{2}\beta^2 + \cdots) \cong Mc^2 + \tfrac{1}{2}Mv^2 + \cdots. \tag{14}$$



The relativistic energy $E = Mc^2/(1 - v^2/c^2)^{1/2}$ and the nonrelativistic kinetic energy $K = \frac{1}{2}Mv^2$ plotted vs v/c. For $v/c \ll 1$, the curves for E and K are almost identical in shape, since $Mc^2/(1 - v^2/c^2)^{1/2} \approx Mc^2 + \frac{1}{2}Mv^2$. For $v/c \sim 1$, E increases much more rapidly than K.

The term $\frac{1}{2}Mv^2$ we recognize as the kinetic energy K in the nonrelativistic limit. Suppose we define the *total relativistic energy* E of a free particle by the equation

$$\boxed{E \equiv Mc^2\gamma \equiv \frac{Mc^2}{(1 - v^2/c^2)^{1/2}}.} \tag{15}$$

Then (12*a*) tells us that

$$\boxed{E^2 - p^2c^2 = M^2c^4,} \tag{16}$$

which is a Lorentz invariant. If we transform from one reference frame to another, with $p \rightarrow p'$ and $E \rightarrow E'$, then the invariance of (16) means that

$$E'^2 - p'^2c^2 = E^2 - p^2c^2 = M^2c^4. \tag{17}$$

This is what we mean when we say that (16) is a Lorentz invariant. We emphasize that M denotes the rest mass of the particle and is a number invariant under a Lorentz transformation.

Transformation of Momentum and Energy. We have seen from (4) and (6) that

$$p_x = M\frac{dx}{d\tau}; \qquad p_y = M\frac{dy}{d\tau}; \qquad p_z = M\frac{dz}{d\tau}. \tag{18}$$

It follows from (4) and (15) that we can write E as

$$E = Mc^2\frac{dt}{d\tau}, \tag{19}$$

using $dt/d\tau = (1 - \beta^2)^{-1/2}$. Because M and τ are Lorentz in variants, it follows from (18) and (19) that p_x, p_y, p_z, and E/c^2 must transform under a Lorentz transformation exactly as x, y, z, and t transform. Because we know how the latter transform, the result (20) follows simply. Using the transformations given in Chap. 11, we have *the transformation relations* for momentum and energy:

$$\boxed{\begin{gathered} p_x' = \gamma\left(p_x - \frac{\beta E}{c}\right); \qquad p_y' = p_y; \qquad p_z' = p_z; \\ E' = \gamma(E - p_x c\beta). \end{gathered}} \tag{20}$$

The inverse transformations follow on changing $-\beta$ into $+\beta$ and interchanging primed and unprimed quantities:

$$\boxed{\begin{gathered} p_x = \gamma\left(p_x' + \frac{\beta E'}{c}\right); \qquad p_y = p_y'; \qquad p_z = p_z'; \\ E = \gamma(E' + p_x' c\beta). \end{gathered}} \tag{20a}$$

We can determine the velocity of the particle from its momentum and energy, using (18) and (19):

$$v_x = \frac{dx}{dt} = \frac{dx}{d\tau}\cdot\frac{d\tau}{dt} = \frac{p_x}{M}\cdot\frac{Mc^2}{E} = \frac{c^2 p_x}{E}, \tag{21}$$

or

$$\boxed{\mathbf{p} = \mathbf{v}\frac{E}{c^2}.} \tag{21a}$$

Example. Inelastic collision. Suppose that two identical particles collide and stick together to make a third particle. In the reference frame S in which the center of mass is at rest we have (by definition of the center of mass):

$$\mathbf{p}_1 + \mathbf{p}_2 = 0; \tag{22}$$

The product particle must be at rest. In another reference frame S′ we have

$$\mathbf{p}_1{}' + \mathbf{p}_2{}' = \mathbf{p}_3{}'. \tag{23}$$

We may express (23) in terms of quantities observed in S by means of the transformation (20):

$$\begin{aligned} p'_{x1} + p'_{x2} &= \gamma(p_{x1} + p_{x2}) - \frac{\gamma\beta(E_1 + E_2)}{c} \\ &= p'_{x3} = \gamma p_{x3} - \frac{\gamma\beta E_3}{c}. \end{aligned} \tag{24}$$

Here E_1 and E_2 are the energies in S of the initial particles; E_3 is the energy in S of the product particle. But $p_{x3} = 0$ and $p_{x1} + p_{x2} = 0$, so that (24) reduces to

$$E_3 = E_1 + E_2. \tag{25}$$

This result tells us that the relativistic energy is conserved in the collision. This discussion will remind you of our discussion of energy and momentum conservation in Chap. 3.

Now because the particles are identical $E_1 = E_2$; using (15) for E_1, E_3 as determined in the frame S, we have

$$M_3c^2 = \frac{2Mc^2}{(1 - v^2/c^2)^{1/2}}. \tag{26}$$

Here M_3 is the *rest* mass of the product particle, and v is the initial velocity of particle 1 or 2 in the frame S. The rest mass M_3 of the product particle is greater in this example than the sum $2M$ of the rest masses of the initial particles. The kinetic energy of the initial particles has been converted into a contribution to the rest mass of the product particle.

In the consideration of the most general collisions it is found that momentum can be conserved only if the sum

$$\sum_i \frac{M_ic^2}{(1 - v_i{}^2/c^2)^{1/2}} = \sum_i E_i \tag{26a}$$

Govt. of West Bengal * STATE INSTITUTE OF EDUCATION *

extended over the incoming particles is equal to the same sum extended over the outgoing particles.† We saw an example of this in (26). That is, momentum can be conserved in a relativistic collision only if the relativistic energy is also conserved.

The new rest mass M_3 is greater than the sum $2M$ of the initial rest masses. For $\beta \ll 1$ we can describe this increase partly in terms of nonrelativistic concepts. Because

$$\frac{1}{(1 - v^2/c^2)^{1/2}} \cong 1 + \frac{v^2}{2c^2} + \cdots, \tag{27}$$

we have from (26) that

$$\begin{aligned} M_3 &\cong 2\left(M + \frac{1}{2}M\frac{v^2}{c^2} + \cdots\right) \\ &\cong 2\left(M + \frac{\text{kinetic energy}}{c^2}\right). \end{aligned} \tag{28}$$

Thus the rest mass M_3 is composed not only of the sum of the rest masses of the incoming particles but also of a contribution proportional to their kinetic energy. In this example of an inelastic collision, (28) shows that there has been a conversion of kinetic energy to mass. [We wrote (28) for small β only because this makes it easier to appreciate the mass-energy conversion. The conversion is valid at high β also.] From (28) the relation between the mass increase

$$\Delta M = M_3 - 2M \tag{29}$$

and the kinetic energy which has disappeared is

$$\text{Kinetic energy} = c^2\,\Delta M. \tag{30}$$

Equivalence of Mass and Energy. The possibility of an interchange between rest mass and energy (and the quantitative relation between them) was considered by Einstein to be the most significant contribution of the theory of relativity. As long as particles never acquire velocities which are significant relative to c, we may use the nonrelativistic definition of kinetic energy, from which we conclude that in any collision among particles (even if the numbers incident and outgoing are different) any net loss or gain in rest mass times c^2 equals the net gain or loss in kinetic energy. Conversely, in an inelastic collision in which there is a loss of kinetic energy, there must be an increase in the rest masses of the outgoing particles.

---

† We cannot apply (26a) directly if photons are involved in the collision, because $v = c$ for a photon. In Eqs. (54) and (55) below we show how to handle the problem for photons and for other particles having zero rest mass.

We see from (9) and (15) that we can write $E = M(v)c^2$. Thus the natural definition of energy in relativity theory is such that the statement (30) is exactly valid for the total energy without restriction to $v/c \ll 1$:

$$\Delta E = \Delta M c^2. \tag{31}$$

(An exact derivation is given in the historical note at the end of Chap. 12.) The mass change ΔM associated with conversion of kinetic energy into rest mass is generally very small in everyday processes, because c is so much greater than common velocities.

Because mass is equivalent to energy, a system with total relativistic energy E has associated with it an inertial mass $M = E/c^2$. Consider a massless box which contains N particles at rest, each of mass M. The box displays an inertial mass NM when we try to accelerate it. The momentum is $NM\mathbf{V}$ if the box is given velocity $\mathbf{V}$. But if each of the particles has a velocity $\mathbf{v}$ and a kinetic energy $\frac{1}{2}Mv^2$ in the box, then the inertial mass of the box is $N(M + Mv^2/2c^2)$ and its momentum is $N\mathbf{V}(M + Mv^2/2c^2)$. In these expressions the velocities V and v have been assumed to be much less than c.

Similarly, a compressed spring has a greater mass than an uncompressed one, by the work needed to compress it divided by c^2. If the compressed spring is completely dissolved in acid, the reaction products are slightly (but too small to measure) more massive than if the spring had not been under stress.

Examples. Mass-energy conversions. (*a*) If two 1-gm masses with equal and opposite velocities of 10^5 cm/sec collide and stick together, the additional rest mass of the joined pair is

$$\Delta M = \frac{\Delta E}{c^2} \cong 2 \cdot \frac{1}{2} M \frac{v^2}{c^2} \cong 1 \times 10^{-11} \text{ gm}. \tag{32}$$

This is less than the precision with which a mass of 1 gm can be measured.

(*b*) A hydrogen atom consists of an electron bound to a proton: its rest mass M_H is lighter than the sum of the rest mass m of the free electron and M_p of the proton. The extra mass of the free particles is equal to the ionization (binding) energy divided by c^2. The mass M_{H} of an H atom is 1.67338×10^{-24} gm. The binding energy of the electron to the proton is known to be 13.6 ev or 22×10^{-12} erg, so that

$$M_p + m - M_{\mathrm{H}} = \frac{22 \times 10^{-12}}{c^2} \cong 2.4 \times 10^{-32} \text{ gm}, \quad (33)$$

which is 1 part in 10^8 of the H atom mass, again too small to be measured.

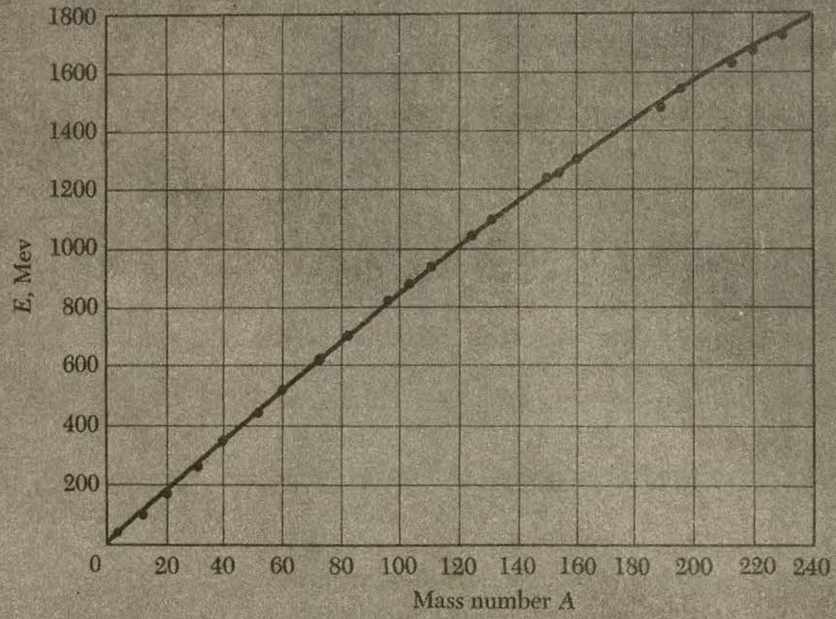
(*c*) The sum of the rest masses of the proton and neutron is,

$$\begin{aligned} M_p + M_n &= (1.6725 + 1.6748) \times 10^{-24} \\ &= 3.3473 \times 10^{-24} \text{ gm}. \end{aligned} \quad (34)$$

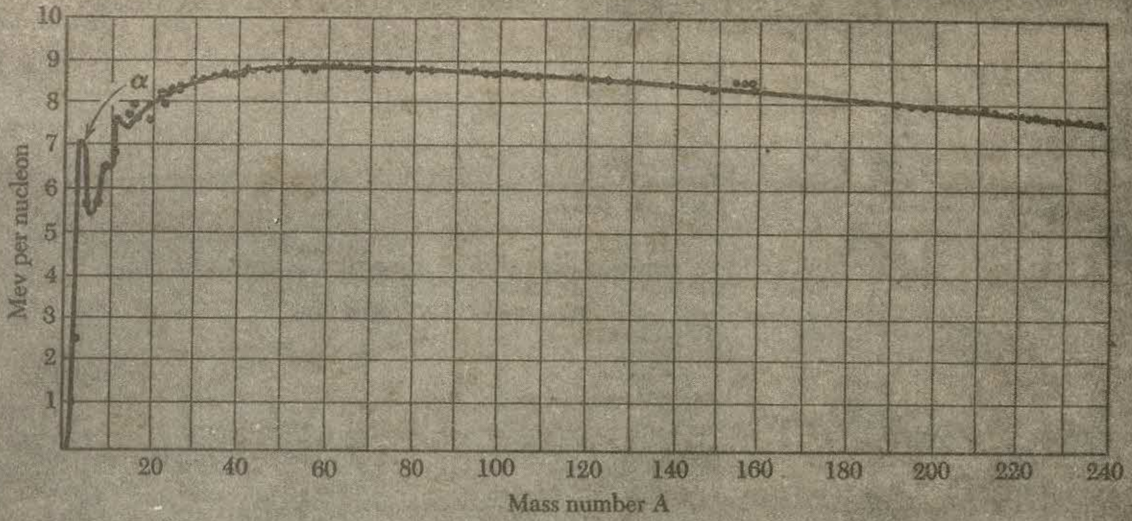
The binding energy of the deuteron with respect to a free proton and neutron is 2.226 Mev, which is equal to

$$\frac{2.226}{0.511} mc^2 \cong 4.36mc^2, \quad (35)$$

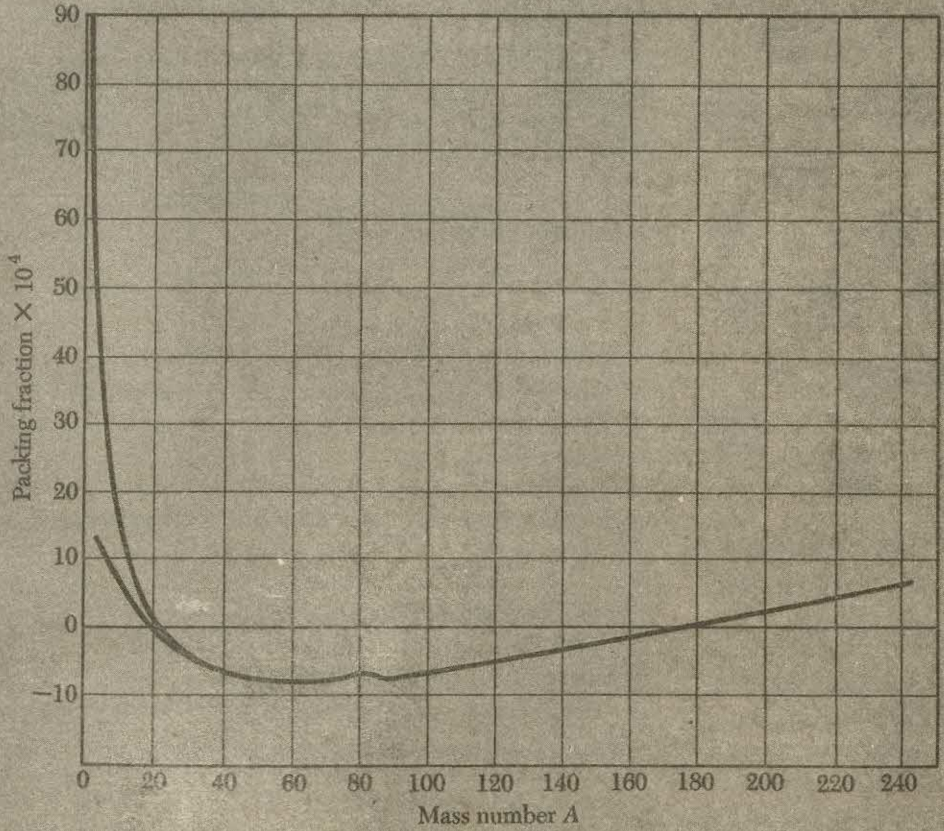
where $m = 0.911 \times 10^{-27}$ gm is the rest mass of the electron.



Binding energy of nuclei, in Mev, as a function of the mass number A of the nucleus. Recall that 1 Mev is equivalent to a mass of 1.76×10^{-27} gm. Not all nuclei are represented in the graph.



The binding energy per nucleon in Mev per nucleon, as a function of the mass number A. The point labeled α corresponds to He^4, which has a relatively large binding energy.



The packing-fraction curve. The mass defect Δ is defined as the difference between the atomic mass M of an isotope and its mass number A: $\Delta = M - A$. The packing fraction F is defined as

$$F = \frac{\Delta}{A} = \frac{M - A}{A}.$$

The observed mass of the deuteron is 3.34334×10^{-24} gm, so that

$$\begin{aligned}\Delta M = M_p + M_n - M_d &= (3.3473 - 3.3433) \times 10^{-24} \\ &= 4.0 \times 10^{-27}\ \text{gm}, \end{aligned} \tag{36}$$

or

$$\Delta M \cong 4.4m; \qquad \Delta Mc^2 = 4.4mc^2, \tag{37}$$

in fair agreement with the value of the binding energy ΔE from (35). We have expressed the result in terms of the electron mass in order to give the reader a feeling for the magnitudes involved. Actually these data are used to obtain the best experimental value of the neutron mass.

(*d*) The accompanying table compares the observed energy release ΔE with the observed mass change ΔM for several nuclear reactions. One unified atomic mass unit (*u*) is equal to one-twelfth the mass of one atom of C^{12}.

***Comparison of calculated and observed energies of disintegration*†**

	Decrease of mass, u	*Energy released, Mev*	
		$\Delta M c^2$	ΔE
$Be^9 + H^1 \rightarrow Li^6 + He^4$	0.00242	2.25	2.28
$Li^6 + H^2 \rightarrow He^4 + He^4$	0.02381	22.17	22.20
$B^{10} + H^2 \rightarrow C^{11} + n^1$	0.00685	6.38	6.08
$N^{14} + H^2 \rightarrow C^{12} + He^4$	0.01436	13.37	13.40
$N^{14} + He^4 \rightarrow O^{17} + H^1$	−0.00124	−1.15	−1.16
$Si^{28} + He^4 \rightarrow P^{31} + H^1$	−0.00242	−2.25	−2.23

† S. Dushman, *General Electric Review* 47, 6–13 (October 1944).

Example. Stellar energy reactions. The most important source of energy in the sun and in most stars arises from the nuclear burning of protons to form helium.

The energy release per helium atom formed can be calculated from the net mass change in the reaction:

$$\begin{aligned} 4M_p + 2m - M(He^4) &= 4(1.6725 \times 10^{-24}) \\ &\quad + 2(0.911 \times 10^{-27}) - 6.647 \times 10^{-24} \\ &\cong 0.045 \times 10^{-24}\ \text{gm} \cong 50m, \end{aligned} \tag{38}$$

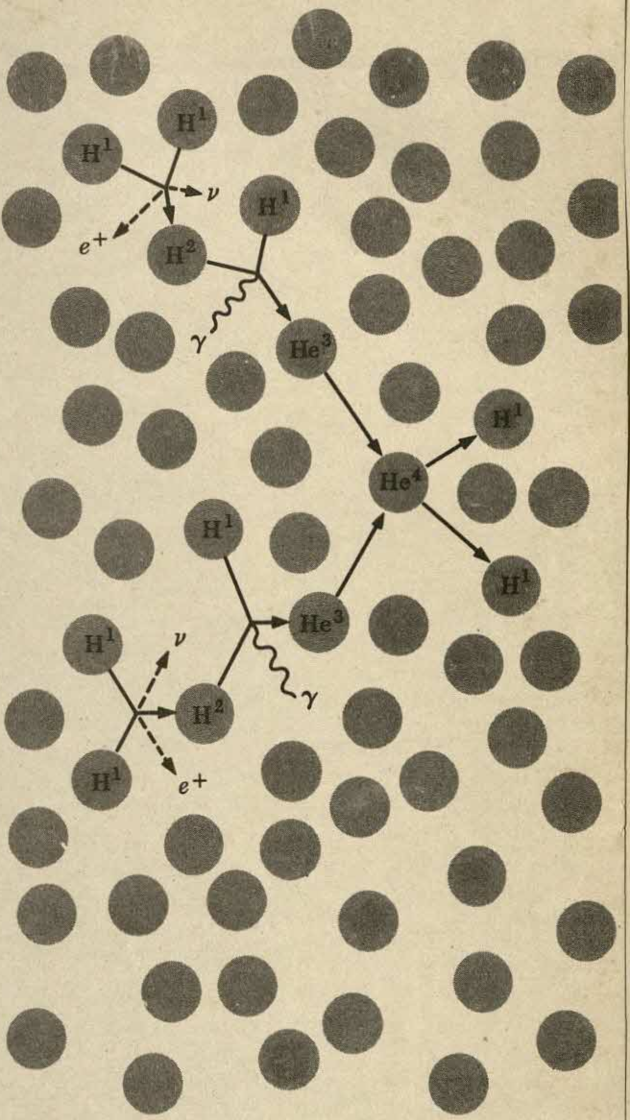
where m denotes the mass of an electron. The result (38) is equivalent to 50(0.511 Mev), or about 25 Mev. The two electrons are included in the reaction to balance the two electrons belonging to an atom of helium. Atomic masses as tabulated usually include the mass of the normal number of atomic electrons; here we have also used the proton mass M_p, which does not include the mass of an electron.

The temperature at the center of the sun is $\sim 2 \times 10^7$ deg K. At this temperature the nuclear processes are believed to be dominated by the following set of reactions:

$$\begin{aligned} \mathrm{H}^1 + p &= \mathrm{H}^2 + e^+ + \text{neutrino}; \\ \mathrm{H}^2 + p &= \mathrm{He}^3 + \gamma; \\ \mathrm{He}^3 + \mathrm{He}^3 &= \mathrm{He}^4 + 2\mathrm{H}^1. \end{aligned}$$

The net effect is to burn hydrogen to produce He^4. Note that a neutrino (massless neutral particle) is given out in the first step, so that the sun is a powerful source of neutrinos. Neutrinos interact very weakly with matter, so that nearly all the neutrinos produced in nuclear reactions in stars escape into space. They may carry off up to 10 percent of the energy emitted from the sun.

For an excellent discussion of the origin of the elements, see William A. Fowler, *Proc. Nat. Acad. Sci.* **52,** 524–548 (1964).



Overall result:
4 hydrogen nuclei → helium nucleus
Energy release = 10^7 kw-hr/lb converted

Schematic representation of the fusion of hydrogen into helium by the *p-p* chain which occurs in main-sequence stars of one solar mass or less. Density: 10^2 gm/cm³. Temperature: 10^7 degrees K. (*After W. A. Fowler*)

Work and Energy. We give here another and more familiar line of argument which supports the definition (15) for the relativistic energy. We saw in Chap. 5 that the rate at which a force **F** does work on a particle is defined by

$$\frac{dW}{dt} = \mathbf{F} \cdot \mathbf{v} = \frac{d\mathbf{p}}{dt} \cdot \mathbf{v}. \tag{39}$$

In the relativistic domain we have

$$\frac{d\mathbf{p}}{dt} = \frac{d}{dt} \frac{M\mathbf{v}}{(1 - v^2/c^2)^{1/2}}, \tag{39a}$$

where M is constant rest mass. By a natural extension of (39) into the relativistic domain we have

$$\frac{dW}{dt} = M\mathbf{v} \cdot \frac{d}{dt} \frac{\mathbf{v}}{(1 - v^2/c^2)^{1/2}}. \tag{40}$$

On a little reflection we see that $\mathbf{v}\cdot\dot{\mathbf{v}} = v\dot{v}$, where

$$\mathbf{v}\cdot\frac{d}{dt}\frac{\mathbf{v}}{(1-v^2/c^2)^{1/2}} = \mathbf{v}\cdot\left\{\frac{\dot{\mathbf{v}}}{(1-v^2/c^2)^{1/2}} + \frac{\mathbf{v}(\dot{v}v/c^2)}{(1-v^2/c^2)^{3/2}}\right\}$$
$$= \frac{1-(v^2/c^2)+(v^2/c^2)}{(1-v^2/c^2)^{3/2}}v\dot{v}$$
$$= \frac{d}{dt}\frac{c^2}{(1-v^2/c^2)^{1/2}}. \tag{41}$$

Here $\dot{v} \equiv dv/dt$. Thus the rate of doing work on the particle is

$$\frac{dW}{dt} = \frac{d}{dt}\frac{Mc^2}{(1-v^2/c^2)^{1/2}}. \tag{42}$$

The natural interpretation of the work done on a free particle is as an increase in the kinetic energy of the particle. On integrating,

$$W = \frac{Mc^2}{(1-v^2/c^2)^{1/2}} + W_0, \tag{43}$$

where W_0 is a constant. Here W is just equal to the kinetic energy K, provided that we take $W = 0$ for the velocity $v = 0$. This determines W_0 as $-Mc^2$, so that the relativistic kinetic energy is

$$K = \frac{Mc^2}{(1-v^2/c^2)^{1/2}} - Mc^2 = Mc^2(\gamma - 1). \tag{44}$$

According to (44) we may write the energy (15) as

$$E = Mc^2 + K. \tag{45}$$

The relativistic energy of a free particle is the sum of the energy associated with the rest mass and of the relativistic kinetic energy For $v/c \ll 1$ we have, as in (27),

$$\frac{1}{(1-v^2/c^2)^{1/2}} \cong 1 + \frac{v^2}{2c^2}. \tag{46}$$

Thus in the limit $v/c \ll 1$ the relativistic kinetic energy (44) is

$$K \cong Mc^2\left(1 + \frac{v^2}{2c^2}\right) - Mc^2 = \frac{1}{2}Mv^2, \tag{47}$$

in agreement in this limit with the nonrelativistic result.

Suppose as in (15) we *define* the total relativistic energy E of a particle as

$$E = \frac{Mc^2}{(1 - v^2/c^2)^{1/2}}. \tag{48}$$

We have seen that, along with the momentum $\mathbf{p}$, this quantity is conserved in a relativistic collision. In the nonrelativistic limit $v/c \ll 1$, the expression for E reduces to

$$E \cong Mc^2 + \tfrac{1}{2}Mv^2 + \cdots. \tag{49}$$

We know from (16) that

$$E^2 = p^2c^2 + M^2c^4. \tag{50}$$

This connects E and p, without solving for the velocity. Part of the energy is associated with the rest mass and part with the momentum. Further, on taking the positive square root,

$$E = (p^2c^2 + M^2c^4)^{1/2} = Mc^2\left[1 + \left(\frac{p^2}{M^2c^2}\right)\right]^{1/2}, \tag{51}$$

which is the relativistic generalization of $E = Mc^2 + \frac{1}{2}Mv^2$.

We have seen in (26*a*) that the conservation of energy assumes the form

$$\sum_{i=1}^{n} E_i = \text{const}, \tag{52}$$

before and after a collision, where E_i is the relativistic energy of the ith particle. The conservation of relativistic energy holds even for what we have called inelastic collisions, because the loss of kinetic energy (into internal excitation of the particles) appears as an increase in the mass. The conservation of momentum assumes the form

$$\sum_{i=1}^{n} \mathbf{p}_i = \text{const}, \tag{53}$$

the same before and after a collision.

Particles with Zero Rest Mass. When $M \equiv 0$ in (50) we have

$$E = pc, \tag{54}$$

so that (21*a*) becomes

$$v = c, \tag{55}$$

and we see that a particle with zero rest mass always moves with the speed of light. It has this same speed for any observer and the same zero rest mass for any observer. Except for the fact that we do not always think of it as a particle, a light pulse in vacuum has just the property $v \equiv c$. In many phenomena in which the quantum nature of light is prominent, we find that light acts as if made up of particles which we call *photons* or light quanta. A photon is a particle of zero rest mass; it is not the only particle of zero rest mass (see Chap. 15). All particles of zero rest mass have the particularly simple property expressed by $E = pc$. The energy of a photon is related to its frequency ν by $E = h\nu$, where h is Planck's constant. Thus $E = h\nu = pc$, or $p = h\nu/c$.

There is always a momentum E/c associated with a photon of energy E. When a photon is absorbed by an atom, a quantity of momentum E/c is transferred to the atom. If the photon is reflected (absorbed and reemitted in the reverse direction), the momentum transfer is $2E/c$.

Let us calculate the pressure of radiation inside a large cube of edge L which contains many photons, with total radiant energy U per unit volume. We assume the photons move in random directions, so that on the average one-third of them moves parallel to any one edge of the cube. In unit time a photon collides on the average with a given cube face $c/6L$ times. The momentum change per collision is $2E/c$. The time-average force on the face is

$$F = (\text{collisions per unit time})(\text{momentum change per collision})$$

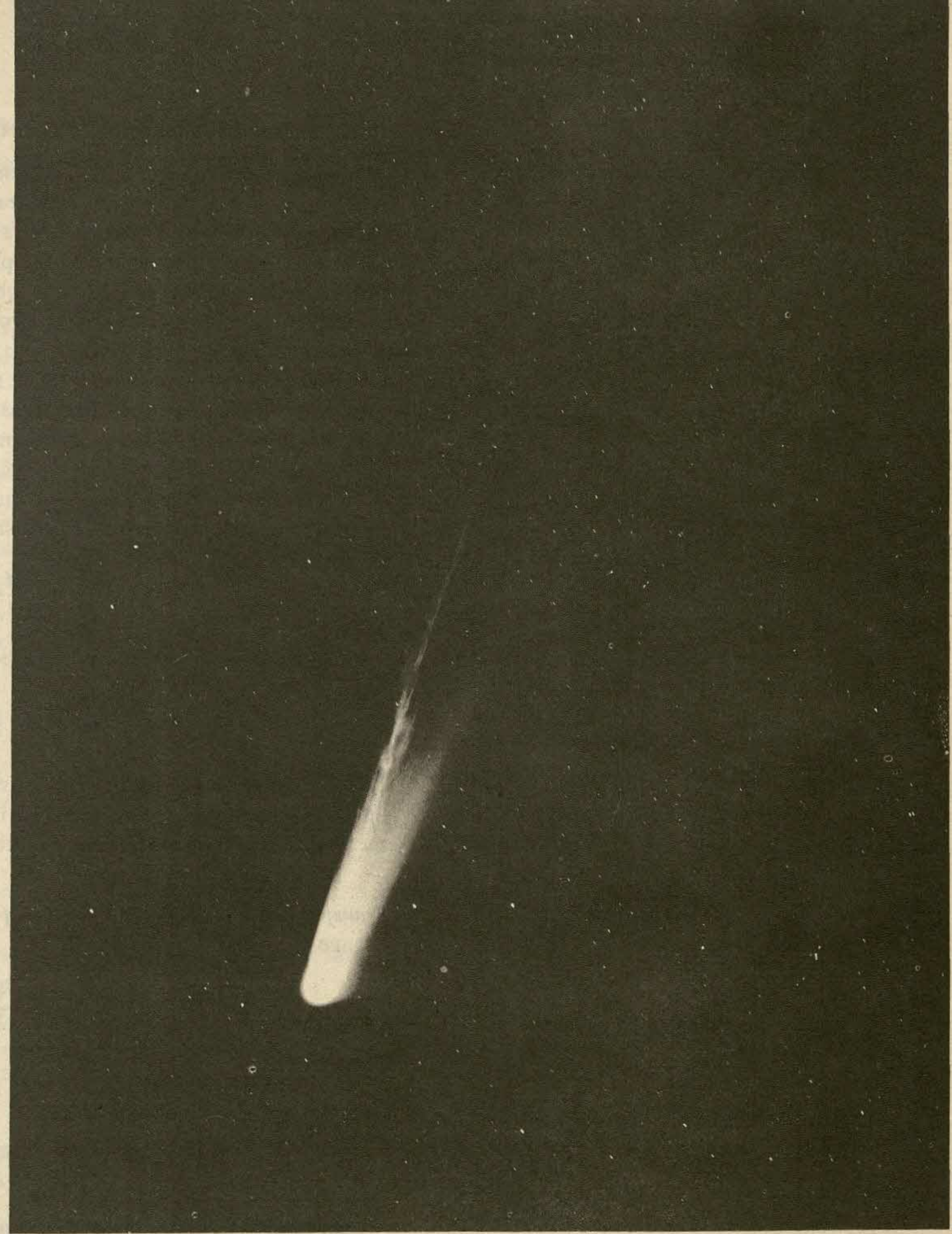
$$= N\left(\frac{c}{6L}\right)\left(\frac{2E}{c}\right) = N\frac{E}{3L}, \qquad (56)$$

where N is the total number of photons in the box. If n is the number of photons per unit volume, then $N = nL^3$, and we have

$$F = nL^2\frac{E}{3}, \qquad \text{or} \qquad P = \frac{1}{3}U, \qquad (57)$$

for the radiation pressure, where $P = F/L^2$ and $U = nE$.

Sunlight deposits on the earth's surface as radiant energy about 10^6 ergs/cm^2-sec. If all the incident energy is absorbed, the resulting pressure is $(10^6/c)$ dynes/cm$^2 \cong 3 \times 10^{-5}$ dynes/cm^2. The pressure is twice as much if the energy is all reflected. This is an extremely small pressure and has an entirely negligible effect on the motion of the earth. On the very diffuse tail of a comet or on Echo satellites, the surface area is large for a given mass and the cumulative effect of such a pressure may not be negligible. Bom-



The Mrkos comet, August 27, 1957. (*Mount Wilson and Palomar Observatories photograph*)

bardment of the comet tail by material particles from the sun may be more important, however. Within the interior of a very hot low-density star the radiation pressure can become enormously significant.

Any particle with sufficiently high energy so that $E \gg Mc^2$ will have approximately the same momentum and energy relations as a photon. For a particle we can always find a moving observer for whom the particle is at rest. But the photon, although its energy and momentum differ for different observers, will always have $v = c = E/p$ and it can never be brought to rest by changing reference frames.

Consider a single hydrogen atom at rest but in an excited electronic state. The hydrogen atom emits a light quantum of energy E and momentum $(E/c)\hat{\mathbf{x}}$. The atom recoils with momentum $-(E/c)\hat{\mathbf{x}}$. In consequence of the recoil, the center of mass of the system (atom plus light quantum) cannot remain at rest unless the light quantum possesses a mass, say M_γ. To find M_γ we set

$$\dot{\mathbf{R}}_{\text{c.m.}} \equiv \frac{M_{\text{H}}\dot{\mathbf{r}}_{\text{H}} + M_\gamma \dot{\mathbf{r}}_\gamma}{M_{\text{H}} + M_\gamma} = 0. \tag{58}$$

Now $M_{\text{H}}\dot{\mathbf{r}}_{\text{H}} = -(E/c)\hat{\mathbf{x}}$ and $\dot{\mathbf{r}}_\gamma = c\hat{\mathbf{x}}$, so that

$$-\frac{E}{c} + M_\gamma c = 0; \qquad M_\gamma = \frac{E}{c^2}. \tag{59}$$

This mass is just that given by the Einstein relation. The mass of the light quantum is not a rest mass; it is the mass equivalent of the energy E. The rest mass of a light quantum is zero.

Transformation of Rate of Change of Momentum. The rate of change of momentum is given by

$$\frac{d\mathbf{p}}{dt} = M\frac{d}{dt}\frac{\mathbf{v}}{(1 - v^2/c^2)^{1/2}}, \tag{60}$$

where M is the rest mass of the particle. In another inertial frame S′ we are concerned with

$$\frac{d\mathbf{p}'}{dt'}.$$

We are trying to avoid the use of the phrase relativistic force (because $d\mathbf{p}/dt$ does not behave as part of a four-vector).

Suppose that the particle is instantaneously at rest in the frame S′, which moves with velocity $v\hat{\mathbf{x}}$ with respect to S. By the Lorentz transformation laws for $\mathbf{p}$ and E, we know that

$$\Delta p_y = \Delta p_y'; \qquad \Delta p_z = \Delta p_z \tag{61}$$

The time interval $\Delta t'$ in S′ is just the proper time interval $\Delta\tau$; hence

$$\Delta t' = \Delta\tau = \left(1 - \frac{v^2}{c^2}\right)^{1/2} \Delta t. \tag{62}$$

It follows that

$$\frac{\Delta p_y}{\Delta t} = \left(1 - \frac{v^2}{c^2}\right)^{1/2} \frac{\Delta p_y'}{\Delta t'} = \left(1 - \frac{v^2}{c^2}\right)^{1/2} \frac{\Delta p_y'}{\Delta\tau}, \tag{63}$$

and similarly for $\Delta p_z/\Delta t$. Thus

$$\boxed{\frac{dp_y}{dt} = \left(1 - \frac{v^2}{c^2}\right)^{1/2} \frac{dp_y'}{d\tau}; \qquad \frac{dp_z}{dt} = \left(1 - \frac{v^2}{c^2}\right)^{1/2} \frac{dp_z'}{d\tau}.} \tag{64}$$

The x components transform differently. We know from (20a) that

$$p_x = \frac{p_x' + vE'/c^2}{(1 - v^2/c^2)^{1/2}}, \tag{65}$$

or

$$\Delta p_x = \frac{\Delta p_x' + v\,\Delta E'/c^2}{(1 - v^2/c^2)^{1/2}}. \tag{66}$$

Now

$$E' = (M^2c^4 + p'^2c^2)^{1/2}, \tag{67}$$

so that

$$\Delta E' = \frac{p_x'\,\Delta p_x' c^2}{(M^2c^4 + p'^2c^2)^{1/2}}. \tag{68}$$

But in the frame S′ we know that $p_x' = 0$ at the instant of time considered; thus (66) becomes

$$\Delta p_x \cong \frac{\Delta p_x'}{(1 - v^2/c^2)^{1/2}}, \tag{69}$$

and, using (62),

$$\frac{\Delta p_x}{\Delta t} \cong \frac{\Delta p_x'}{\Delta t'} = \frac{\Delta p_x'}{\Delta\tau}. \tag{70}$$

In the limit $\Delta t \to 0$ we have

$$\boxed{\frac{dp_x}{dt} = \frac{dp_x'}{d\tau}.} \tag{71}$$

The results (64) and (71) play an important part in the development of electromagnetism in Vol. II.

Constancy of Charge. The law of motion $q\mathbf{E} = \dot{\mathbf{p}}$ of a particle of charge q in an electric field $\mathbf{E}$ is incomplete unless we know the dependence of the charge on the speed and acceleration of the particle whose momentum is $\mathbf{p}$. The best experimental evidence that the charge on a proton or electron is very accurately constant is the observation that beams of hydrogen atoms or molecules suffer no deflection in uniform electric fields perpendicular to the beam. The hydrogen atom consists of an electron (e) and a proton (p). The H_2 molecule consists of two electrons and two protons. Even when the protons are moving very slowly, the electrons move around the protons with an average velocity of about $10^{-2}c$. An undeviated molecule has constant momentum, so that the experimental result tells us that $\dot{\mathbf{p}}_p + \dot{\mathbf{p}}_e = 0 = (e_p + e_e)\mathbf{E}$. Thus it follows from experiment that in the atom or molecule $e_e = -e_p$, despite the fact that the electron has a high velocity, and even though the average velocity of the electron differs in the atom and the molecule. Quantitatively, the electron charge is known to be independent of velocity and equal to that of the proton to at least one part in 10^9 up to an electron velocity of $10^{-2}c$.

The experimental situation is discussed in Vol. II. The experimental result is that the charge is observed to be independent of the velocity of the particle or the observer. Thus charge and mass transform in different ways when the frame of reference is changed.

Problems

1. *Relativistic momentum.* What is the momentum of a proton having kinetic energy† 1 Bev? (If E is measured in Bev, we may measure p in Bev/c.) *Ans.* 1.7 Bev/c.

2. *Relativistic momentum.* What is the momentum of an electron having kinetic energy 1 Bev? *Ans.* 1.0005 Bev/c.

3. *Photon momentum.* What is the momentum of a photon of energy 1 Bev?

4. *Energy and momentum of fast proton.* Given a proton whose $\beta = 0.995$ measured in the laboratory; what are the corresponding relativistic energy and momentum?

5. *Energetic cosmic-ray particles.* It is known that cosmic-ray particles have energies up to 10^{19} ev, and perhaps higher.

(*a*) What is the apparent mass of such a particle (approximately)? *Ans.* 1.8×10^{-14} gm.

---

† The word *billion* has different meanings in different countries. In the United States it means 10^9, but in British usage it means 10^{12}. The prefix *giga* (abbreviated G) is often used to denote 10^9. Thus 1 Bev (American) $\equiv$ 1 Gev $\equiv$ 10^9 ev.

(*b*) What is the momentum (approximately)?

Ans. 5×10^{-4} gm-cm/sec.

6. *Transformation of energy and momentum.* A proton has $\beta = 0.999$ in the laboratory. Find the energy and momentum as observed in a frame traveling in the same direction, with $\beta' = 0.990$ with respect to the laboratory.

7. *Energy of fast electron.* An electron has $\beta = 0.99$. What is its kinetic energy? *Ans.* 3.1 Mev.

8. *Recoil in γ-ray emission.* What is the recoil momentum in the laboratory of an Fe^{57} nucleus recoiling due to the emission of a 14-kev photon? Is the momentum of the nucleus relativistic? *Ans.* 7.5×10^{-19} gm-cm/sec.

9. *Recoil of nucleus.* Consider a gamma ray of energy E_γ which strikes a proton at rest in the laboratory.

(*a*) In the laboratory frame what is the momentum of the gamma ray?

(*b*) Show that the velocity V of the center of mass in the laboratory frame is given by

$$\frac{V}{c} = \frac{E_\gamma}{E_\gamma + M_p c^2}.$$

10. *Neutron decay.* Use values given in Chap. 12 to calculate the amount of energy released as a neutron decays into a proton and an electron.

Ans. 0.79 Mev.

11. *Lorentz invariance in two-particle system.* Let the total momentum and energy for a two-particle system be $\mathbf{p} = \mathbf{p}_1 + \mathbf{p}_2$ and $E = E_1 + E_2$, respectively. Show explicitly that the Lorentz transformations on $\mathbf{p}$ and E are consistent with the invariance of the quantity $E^2 - p^2c^2$.

12. *Transformation to rest frame from center-of-mass reference frame.* Two protons travel in opposite directions from a common point with velocities $\beta = 0.5$.

(*a*) What are the energy and momentum of one proton relative to the common point?

(*b*) Use the Lorentz transformation to find the energy and momentum of one proton in the rest frame of the other. (In problems of this sort, it is usually convenient to express the energy as a multiple of some rest mass energy.)

13. *Radiation of mass by radio transmitter.* What is the mass equivalent of the energy from an antenna radiating 1000 watts of radio energy for 24 hours? (1 watt $\equiv 10^7$ ergs/sec.)

14. *Solar energy.* The *solar constant* is the flux of solar energy per square centimeter per second at the distance of the earth from the sun. By measurement it is found that the value of the constant is 1.4×10^6 ergs/sec-cm^2.

(*a*) Show that the total energy generation of the sun is $\approx 4 \times 10^{33}$ ergs/sec.

(*b*) Show that the average rate of energy generation per gram of matter on the sun is ≈ 2 ergs/gm-sec $\approx 6 \times 10^7$ ergs/gm-yr.

(*c*) Using (38), show that the energy equivalent of 1 gm of hydrogen burned to produce He^4 is $\approx 6 \times 10^{18}$ ergs.

(*d*) Show that if the mass of the sun were one-third hydrogen and the nuclear burning process continued without change, then the sun could continue to radiate at its present rate for 3×10^{10} years.

15. *Radiation propulsion.* One possible means of propulsion in space is a large reflecting metallic sheet connected to a small vehicle. Make reasonable estimates of the accelerations which might result for some typical vehicle at a distance of 1 A.U. from the sun.

16. *Momentum of laser pulse.* A large laser can produce a pulse of light having an energy of 2000 joules. (1 joule $\equiv 10^7$ ergs.)

(*a*) Show that the momentum of the pulse is of the order of 1 gm-cm/sec.

(*b*) Discuss how you might detect this momentum. The duration of the pulse might be 1 millisecond (10^{-3} sec).

Historical Note. Mass-energy relation

Einstein's first paper on the special theory of relativity, titled "On the electrodynamics of moving bodies," appeared in *Annalen der Physik* 17, 891–921 (1905). This volume of the *Annalen* contains three classic papers by Einstein: one on the quantum interpretation of the photoelectric effect (pp. 132–148); one on the theory of Brownian motion (pp. 549–560); and that on special relativity cited above (many of the results in this paper had been anticipated by Larmor, Lorentz, and others). In the same year a short paper by Einstein appeared in vol. 18, pp. 639–641; this was entitled "Does the inertia of a body depend on its energy content?" We paraphrase here Einstein's argument:

Consider (as in Einstein's paper on electrodynamics) a packet or group of plane waves of light. Let the packet possess the energy ϵ and move in the positive x direction in the frame S. As viewed in the frame S′, which moves with velocity $V\hat{\mathbf{x}}$ relative to S, the wave packet has the energy

$$\epsilon' = \epsilon\left(\frac{1-\beta}{1+\beta}\right)^{1/2}; \qquad \beta = \frac{V}{c}. \tag{72}$$

This result was derived in Einstein's paper on electrodynamics, without reference to the notion of a photon. It follows most directly from another argument: We note from the result (11.41) of the longitudinal Doppler effect that the frequencies seen by observers at rest in S′ and S are related by

$$\nu' = \nu\left(\frac{1-\beta}{1+\beta}\right)^{1/2}. \tag{73}$$

According to the quantum picture a light pulse may be viewed as made up of an integral number of light quanta or *photons*, each of energy $h\nu$ (as viewed in S), where h is Planck's constant. When we view the pulse from S′, the number of photons remains unchanged, but the energy of a photon becomes $h\nu'$. (We suppose that the value of h is the same in S′ as in S.) If follows that the energy ϵ' of the light pulse is proportional to ν', whence (72) follows from (73).

Now let there be a stationary body in the system S, and let its initial energy be E_0 in S and E_0' in S′. We suppose that the body emits a light pulse of energy $\frac{1}{2}\epsilon$ in the positive x direction and a similar pulse of energy $-\frac{1}{2}\epsilon$ in the negative x direction. The body will remain at rest in S. Let E_1, E_1' denote in S, S′ the energy of the body after emission of the two light pulses. Then by conservation of energy,

$$E_0 = E_1 + \frac{1}{2}\epsilon + \frac{1}{2}\epsilon; \tag{74}$$

$$\begin{aligned} E_0' &= E_1' + \frac{1}{2}\epsilon\left(\frac{1-\beta}{1+\beta}\right)^{1/2} + \frac{1}{2}\epsilon\left(\frac{1+\beta}{1-\beta}\right)^{1/2} \\ &= E_1' + \frac{\epsilon}{(1-\beta^2)^{1/2}}, \end{aligned} \tag{75}$$

from which, on subtraction of (75) from (74), we have

$$E_0 - E_0' = E_1 - E_1' + \epsilon - \frac{\epsilon}{(1-\beta^2)^{1/2}}. \tag{76}$$

Now the difference in energy $E_0' - E_0$ must be just the initial kinetic energy K_0 of the body as viewed in S′, because the body is initially at rest in S. Similarly, $E_1' - E_1$ is the final kinetic energy K_1 as viewed in S′. Thus (76) may be written as

$$K_0 - K_1 = \epsilon\left(\frac{1}{(1-\beta^2)^{1/2}} - 1\right). \tag{77}$$

We see that the kinetic energy of the body diminishes as a result of the emission of light. The amount of diminution is independent of the properties of the body. If $\beta \ll 1$,

$$K_0 - K_1 \cong \frac{1}{2}\epsilon\beta^2 = \frac{1}{2}\frac{\epsilon}{c^2}V^2, \tag{78}$$

so that the rest mass of the body decreases by

$$\Delta M = \frac{\epsilon}{c^2}. \tag{79}$$

From this relation Einstein concluded:

"*If a body gives off the energy ϵ in the form of radiation, its mass diminishes by ϵ/c^2.* The fact that the energy withdrawn from the body becomes energy of radiation evidently makes no difference, so that we are led to the more general conclusion that

"The mass of a body is a measure of its energy content; if the energy changes by ϵ, the mass changes in the same sense by $\epsilon/(9 \times 10^{20})$, the energy being measured in ergs, and the mass in grams.

"It is not impossible that with bodies whose energy content is variable to a high degree (e.g. with radium salts), the theory may be successfully put to the test.

"If the theory corresponds to the facts, radiation conveys inertia between the emitting and absorbing bodies."

Chapter 13

Acceleration of charged particle by constant longitudinal electric field 408

Example: Acceleration by a transverse electric field 410

Charged particle in magnetic field 411

Center-of-mass system and threshold energy 413

Examples: Gamma-ray recoil 415

Energy available from moving particle 415

Antiproton threshold 417

Relativistic rocket equation 418

Suggested reading 419

Film list 419

Problems 420

Historical Note: Synchrotron 422

Simple Problems in Relativistic Dynamics

In Chap. 4 we discussed a number of problems involving the nonrelativistic motion of particles in electric and magnetic fields. In Chap. 3 and again in Chap. 6, we discussed elastic and inelastic collisions of two nonrelativistic particles. Now we extend several of the earlier solutions to the relativistic region. Often the solutions present no special difficulty, and several of them are of the highest importance to the worlds of high-energy particle physics and to astrophysics.

The fact that the momentum of an accelerated relativistic particle can increase indefinitely, even when the velocity approaches very close to the velocity of light, forms the basis for the great accelerators and for the momentum analysis of high-energy particles by means of a deflecting magnetic field. Magnetic deflection methods are widely used in cosmic-ray research and in other work with high-energy particles.

The threshold energy of a relativistic particle reaction is much larger when one looks at the reaction in the laboratory system compared to the center-of-mass system. This effect is a major handicap on the frontiers of research in particle physics.

We treat first the deflection of a relativistic particle by an electric field in order to gain familiarity with the use of certain standard manipulations.

Acceleration of Charged Particle by Constant Longitudinal Electric Field. The equation of motion of a particle of charge q and rest mass M in a uniform constant electric field† $\mathcal{E}\hat{\mathbf{x}}$ is

$$\dot{p}\,\hat{\mathbf{x}} = q\mathcal{E}\hat{\mathbf{x}}, \tag{1}$$

or, with $\mathbf{p} = M\mathbf{v}/(1 - v^2/c^2)^{1/2}$,

$$M\frac{d}{dt}\frac{v}{(1 - v^2/c^2)^{1/2}} = q\mathcal{E}, \tag{2}$$

if we assume $v_y = v_z = 0$, as required for acceleration from rest in the x direction. On integrating (2) with respect to the time, we see directly that

$$M\frac{v}{(1 - v^2/c^2)^{1/2}} = q\mathcal{E}t, \tag{3}$$

with $v(0) = 0$. After squaring both sides and rearranging, we obtain

$$v^2 = \frac{(q\mathcal{E}t/Mc)^2}{1 + (q\mathcal{E}t/Mc)^2}c^2. \tag{4}$$

For short times‡ $t < Mc/q\mathcal{E}$, the denominator in (4) may be replaced by unity and we have

$$v^2 \cong \left(\frac{q\mathcal{E}}{M}t\right)^2, \tag{5}$$

just as in the nonrelativistic approximation in Chap. 4.

For long times $t \gg Mc/q\mathcal{E}$, we have the relativistic approximation

$$v^2 \cong \left[1 - \left(\frac{Mc}{q\mathcal{E}t}\right)^2\right]c^2, \tag{6}$$

which shows how v approaches c as a limiting velocity. In the long time limit we have

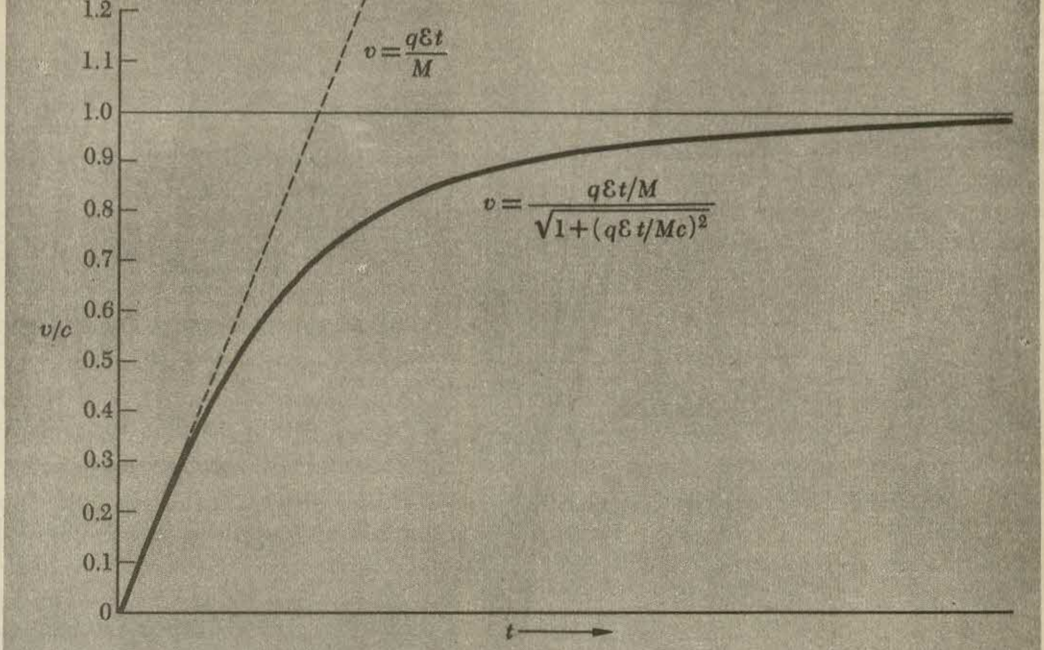
$$\frac{v^2}{c^2} \cong 1 - \left(\frac{Mc}{q\mathcal{E}t}\right)^2, \tag{7}$$

---

† Here we denote the electric field intensity by $\mathcal{E}$, to avoid confusion with the energy E.

‡ With $\mathcal{E} = 1$ statvolt/cm, we have for an electron

$$\frac{mc}{e\mathcal{E}} \approx \frac{(10^{-27})(3 \times 10^{10})}{(5 \times 10^{-10})} \approx 10^{-7}\text{ sec.}$$



The velocity v of a charge q of rest mass M, accelerated from rest by a uniform electric field $\mathcal{E}$, is plotted versus time. The velocity v approaches the limit c for $t \gg 0$. The dotted line represents the velocity of the charge as predicted by nonrelativistic mechanics.

and, with $\beta \equiv v/c$,

$$\frac{1}{(1-\beta^2)^{1/2}} \cong \frac{q\mathcal{E}t}{Mc}. \tag{8}$$

From Chap. 12 and Eq. (8) above, the relativistic energy is given by

$$E = \frac{Mc^2}{(1-\beta^2)^{1/2}} \approx q\mathcal{E}ct, \tag{9}$$

again in the long time limit $t \gg Mc/q\mathcal{E}$. The limiting result is just the force times the distance traveled in time t at velocity c. In the same limit the momentum is

$$p \approx \frac{Mc}{(1-\beta^2)^{1/2}} \approx q\mathcal{E}t. \tag{10}$$

Note the characteristic relativistic feature that the momentum may go on increasing even after the velocity has for all practical purposes leveled off.

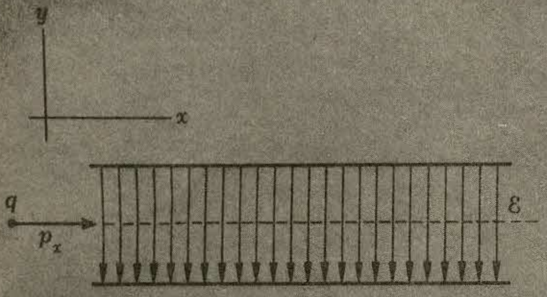
The displacement x is found from the square root of (4), writing dx/dt for v, and s for $q\mathcal{E}/Mc$:

$$dx = \frac{cst}{(1 + s^2t^2)^{1/2}}\,dt. \tag{11}$$

After integrating between 0 and t, we obtain the displacement

$$x = \frac{Mc^2}{q\mathcal{E}}\left\{\left[1 + \left(\frac{q\mathcal{E}t}{Mc}\right)^2\right]^{1/2} - 1\right\}. \tag{12}$$

(Check the algebra for yourself.) We have supposed that $x(0) = 0$ and $v(t) = 0$. At long times $x \approx ct$, as if the particle moved nearly at the speed of light.



Suppose a charge q with initial momentum p_x enters a transverse field ℰ.

Example. Acceleration by a transverse electric field. We consider a charged particle which moves along the x axis with a high momentum p_0 and enters a region of length L in which there is a transverse electric field $\mathcal{E}\hat{\mathbf{y}}$. Find the angle through which the particle is deflected by the electric field.

The equations of motion are

$$\frac{dp_x}{dt} = 0; \qquad \frac{dp_y}{dt} = q\mathcal{E}, \tag{13}$$

from which

$$p_x = p_0; \qquad p_y = q\mathcal{E}t. \tag{14}$$

We want to find the velocity $\mathbf{v}$. If we can find the energy E, then we can find the velocity from the momentum by using the relation $\mathbf{v} = \mathbf{p}c^2/E$ derived in Chap. 12.

The energy is given by

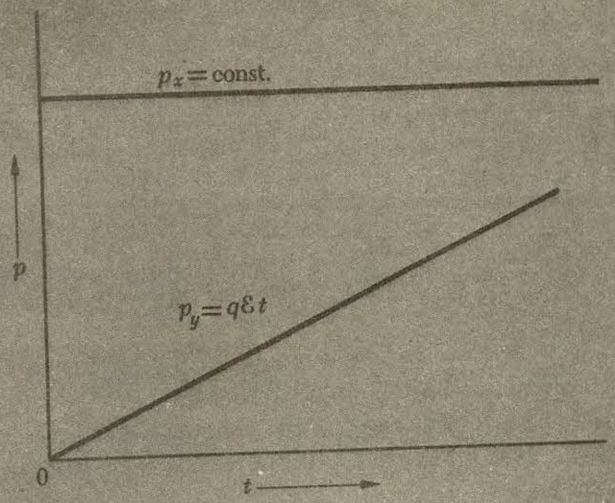
$$\begin{aligned} E^2 = M^2c^4 + p^2c^2 &= M^2c^4 + p_0{}^2c^2 + (q\mathcal{E}tc)^2 \\ &= E_0{}^2 + (q\mathcal{E}tc)^2, \end{aligned} \tag{15}$$

where E_0 is the initial energy. Therefore from (14) and the velocity-momentum relation we have

$$v_x = \frac{p_0c^2}{[E_0{}^2 + (q\mathcal{E}tc)^2]^{1/2}}; \tag{16}$$

$$v_y = \frac{q\mathcal{E}tc^2}{[E_0{}^2 + (q\mathcal{E}tc)^2]^{1/2}}. \tag{17}$$

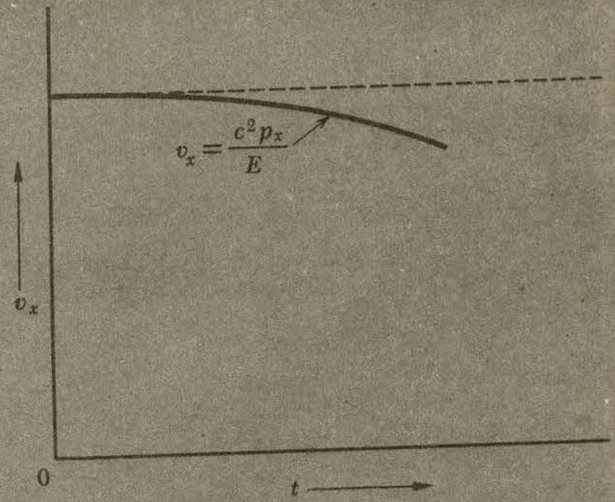
Note that v_x decreases as t increases. We see also that v_y is



The force in the y direction is $q\mathcal{E}$, so $p_y = q\mathcal{E}t$, while p_x remains constant. The energy

$$E = c\sqrt{(p_x^2 + p_y^2) + M^2c^2}$$

increases.



Since $v_x = c^2p_x/E$, v_x actually *decreases* when the particle is accelerated in the y direction. Nonrelativistic mechanics would, of course, predict $v_x = \text{const}$.

always less than the nonrelativistic value $q\mathcal{E}t/M$. At a time t, the angle θ the trajectory makes with the x axis is given by

$$\tan\theta(t) = \frac{v_y}{v_x} = \frac{q\mathcal{E}tc^2}{p_0c^2} = \frac{q\mathcal{E}t}{p_0}. \tag{18}$$

The time t_L required to traverse the distance L is found on solving (16) for x:

$$\int_0^L dx = p_0c^2 \int_0^{t_L} \frac{dt}{[E_0^2 + (q\mathcal{E}tc)^2]^{1/2}}, \tag{19}$$

or, by Dwight 728.1,

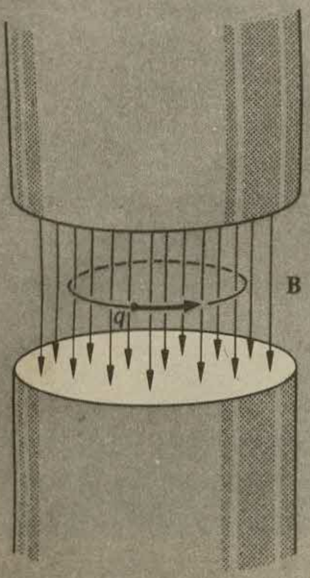
$$L = \frac{p_0c}{q\mathcal{E}} \sinh^{-1}\left(\frac{q\mathcal{E}t_Lc}{E_0}\right), \tag{20}$$

so that

$$t_L = \frac{E_0}{q\mathcal{E}c} \sinh \frac{q\mathcal{E}L}{p_0c}. \tag{20a}$$

Charged Particle in Magnetic Field. We consider an important practical problem dealing with the motion of a particle of charge q in a uniform constant magnetic field $\mathbf{B}$. The equation of motion is

$$\frac{d\mathbf{p}}{dt} = \frac{q}{c}\mathbf{v} \times \mathbf{B}. \tag{21}$$



A charge with velocity **v** perpendicular to a uniform magnetic field describes circular motion, with radius $\rho = pc/qB$.

Just as in the nonrelativistic problem (Chap. 4), we have $dp^2/dt = 0$, because

$$\frac{d}{dt}p^2 = 2\mathbf{p}\cdot\frac{d\mathbf{p}}{dt} = 2\frac{q}{c}\mathbf{p}\cdot\mathbf{v}\times\mathbf{B}; \tag{22}$$

but **p** is always parallel to **v**, so that the triple product is zero. Thus the magnitude of the momentum and consequently the magnitude of the velocity of the particle are unchanged by a constant magnetic field. But if only the direction is altered by the field, then the factor

$$\frac{M}{(1 - v^2/c^2)^{1/2}}, \tag{23}$$

which enters into the definition of **p**, is constant.

The equation of motion (21) can now be written as

$$\frac{d\mathbf{p}}{dt} = \frac{M}{(1 - v^2/c^2)^{1/2}}\frac{d\mathbf{v}}{dt} = \frac{q}{c}\mathbf{v}\times\mathbf{B}. \tag{24}$$

Because of the constancy of (23) this equation has solutions in which the particle moves in a circle. Let ρ denote the radius of the circle and ω_c the angular frequency of the motion. On substituting in (24) the centripetal acceleration $\omega_c{}^2\rho$ for $d\mathbf{v}/dt$ and $\omega_c\rho$ for v, we have

$$\frac{M}{(1 - v^2/c^2)^{1/2}}\,\omega_c{}^2\rho = \frac{q}{c}\,\omega_c\rho B, \tag{25}$$

from which

$$\boxed{\omega_c = \frac{qB(1 - v^2/c^2)^{1/2}}{Mc}.} \tag{26}$$

We see that the frequency of the motion is lower for fast particles than for slow particles. Thus a cyclotron can be used to accelerate particles to relativistic energies only if the frequency of the rf accelerating field (or the magnetic field intensity) is modulated to remain synchronous with (26) as the energy of the particles is increased. For nonrelativistic particles the dependence of the frequency on the velocity may be neglected.

Values for ω_c predicted from (26) have been confirmed experimentally in the operation of high-energy accelerators. The relation has been confirmed for electrons accelerated in a synchrotron in situations where $1/(1 - \beta^2)^{1/2} \approx 2000$; that is, when the apparent mass of the particle is 2000 times the rest mass. It is interesting to see what this means in terms of $c - v$, the amount by which the speed of light exceeds the speed of the particle. We have

$$(1 - \beta^2) = (1 + \beta)(1 - \beta) \cong 2(1 - \beta) \cong (2000)^{-2}. \qquad (27)$$

Note that we have set $(1 + \beta) \cong 2$. From (27) we have

$$1 - \beta = \frac{c - v}{c} \approx 10^{-7}; \qquad c - v \approx 3 \times 10^3 \text{ cm/sec}. \qquad (28)$$

It is believed that protons in cosmic radiation have been detected with $c - v \approx 10^{-12}$ cm/sec; here the value of β is inferred from the energy observed to be dissipated by the particle in collision events in the atmosphere.

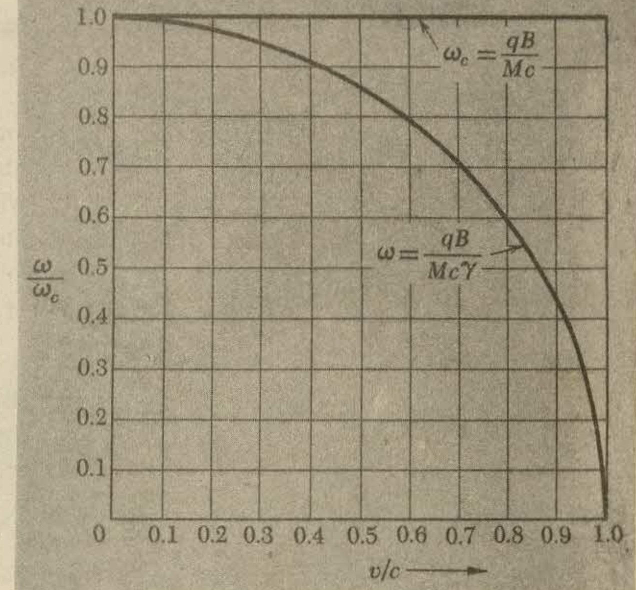
Using (26), the gyroradius ρ of a relativistic particle in a magnetic field is given by

$$\rho = \frac{v}{\omega_c} = \frac{cMv}{qB(1 - \beta^2)^{1/2}}. \qquad (29)$$

But the right-hand side contains the momentum p, so that

$$\boxed{B\rho = \frac{cp}{q}.} \qquad (30)$$

Therefore the radius ρ of the circle traced out by a charged particle in a magnetic field is a direct measure of the relativistic momentum. This relation is the basis of the most important single method for measuring the momentum of charged relativistic particles. It is used in the bubble-chamber photographs given in Chap. 15.



The cyclotron frequency ω for circular motion of a charge q, rest mass M, in a plane perpendicular to a uniform magnetic field **B**, is plotted versus the velocity ratio v/c. The nonrelativistic cyclotron frequency ω_c is represented by the horizontal line.

Center-of-mass System and Threshold Energy. Conservation of energy imposes a general limitation on the nuclear reactions or events which can take place when two particles collide. For example, a high-energy photon (γ ray) can produce an electron-positron pair according to the reaction

$$\gamma \rightarrow e^- + e^+ \qquad (31)$$

only if the energy of the γ ray exceeeds the energy equivalent of the rest masses of the electron and positron. Thus conservation of energy alone dictates that the threshold or minimum energy for the production of an electron-positron pair is

$$E_\gamma = 2mc^2 \cong 1.02 \times 10^6 \text{ ev}. \qquad (32)$$

We recall that the rest mass of the positron is equal to the rest mass of the electron.

This reaction, however, is impossible in free space at any energy, because momentum cannot be conserved. We saw in Chap. 12 that the momentum of the photon is $p_\gamma = E_\gamma/c$. We choose to view the reaction in the reference frame in which the center of mass of the electron-positron pair is at rest. In this frame the sum of the electron momentum and positron momentum is zero:

$$\mathbf{p}_{e^-} + \mathbf{p}_{e^+} = 0. \tag{33}$$

But in this frame the momentum of the incident photon is not zero, because there is *no* reference frame in which the momentum of a photon can be made to vanish.† Therefore in the center-of-mass frame

$$\mathbf{p}_\gamma \neq \mathbf{p}_{e^-} + \mathbf{p}_{e^+} = 0, \tag{34}$$

and the reaction (31) is not correct (does not take place) because momentum is not conserved. If it cannot take place in one reference frame, it cannot take place in any reference frame.

The reaction can proceed in the vicinity of another particle, such as a nucleus of an atom, for then the nucleus can absorb the momentum change. It absorbs the change by pushing with its Coulomb field on the charged particle. We can have

$$\mathbf{p}_\gamma + \mathbf{p}_{\text{nuc}} = \mathbf{p}'_{\text{nuc}} + \mathbf{p}_{e^-} + \mathbf{p}_{e^+}. \tag{35}$$

The nucleus has its momentum changed by the reaction, but otherwise the nucleus is unchanged and acts only as a catalyst of a very simple kind. The initial nuclear momentum may be zero.

A heavy particle or nucleus is a good vehicle for absorbing excess momentum without absorbing much energy. We see this from the form of the nonrelativistic kinetic energy

$$K = \frac{1}{2}Mv^2 = \frac{p^2}{2M}. \tag{36}$$

The larger the mass M, the smaller the kinetic energy associated with a given momentum.

---

† We change the frequency of a photon by changing reference frames, but we can neither make it disappear nor bring it to rest in this way.

Example. Gamma-ray recoil. A nucleus of mass M emits a γ ray of energy E_γ. The nucleus was initially at rest. What is its recoil energy after the emission?

From momentum conservation

$$0 = \mathbf{p}_\gamma + \mathbf{p}_{\text{nuc}}, \tag{37}$$

where $p_\gamma = E_\gamma/c$. Thus the recoil momentum is

$$|\mathbf{p}_{\text{nuc}}| = \frac{E_\gamma}{c}, \tag{38}$$

and the recoil energy is

$$K_{\text{nuc}} = \frac{p_{\text{nuc}}{}^2}{2M} = \frac{E_\gamma{}^2}{2Mc^2}. \tag{39}$$

We have assumed that the recoil velocity may be treated as nonrelativistic.

In collision events in which new particles are created, the requirement of momentum conservation usually makes it impossible to convert all the initial kinetic energy in the laboratory system into rest mass of new particles formed in the collision. If there is a net momentum in the initial state before the collision, there must be an equal momentum in the final state after the collision. Therefore the particles remaining after the collision will not be at rest; some of the initial kinetic energy is transferred as kinetic energy of the final particles.

The only situation in which *all* the initial kinetic energy is available for the reaction occurs when the momentum of the initial state is zero. The momentum can always be made to appear zero by viewing the collision from a suitable reference frame, the center-of-mass reference frame.

Example. Energy available from moving particle. How much energy is available in the collision of a moving proton with a proton at rest?

Suppose first that the kinetic energy of the incident proton is much less than M_pc^2, so that the collision may be treated nonrelativistically. If the incident proton has velocity $\mathbf{v}$ in the laboratory reference frame, its kinetic energy is

$$K_{\text{lab}} = \tfrac{1}{2}M_pv^2. \tag{40}$$

In the center-of-mass reference frame one proton has velocity $\frac{1}{2}\mathbf{v}$ and the other has velocity $-\frac{1}{2}\mathbf{v}$. In the center-of-mass reference frame all the kinetic energy is available for the production of further particles; this kinetic energy is

$$K_{\text{c.m.}} = \tfrac{1}{2}M_p(\tfrac{1}{2}v)^2 + \tfrac{1}{2}M_p(\tfrac{1}{2}v)^2 = \tfrac{1}{4}M_pv^2. \tag{41}$$

From (40) and (41) we have the nonrelativistic result

$$\frac{K_{\text{c.m.}}}{K_{\text{lab}}} = \frac{1}{2}; \tag{42}$$

thus one-half of the energy in the laboratory system is available. If we accelerate a proton to 200 Mev, only 100 Mev of this energy may be used in a collision with another proton to create other particles.

The efficiency is lower in the relativistic region. It is simple and helpful to carry out the calculation for the extreme relativistic region for which the kinetic energy of a proton even in the center-of-mass system is much larger than the rest energy M_pc^2.

We can relate the total relativistic energy in the laboratory system to the total relativistic energy in the center-of-mass system by using the invariance property (12.16) applied to the system of two protons:

$$\underbrace{(E_1 + E_2)^2 - (\mathbf{p}_1 + \mathbf{p}_2)^2c^2}_{\text{lab}} = \underbrace{(E_1 + E_2)^2 - (\mathbf{p}_1 + \mathbf{p}_2)^2c^2}_{\text{c.m.}}. \tag{43}$$

By definition of the center-of-mass system $(\mathbf{p}_1 + \mathbf{p}_2)_{\text{c.m.}} = 0$. If proton 2 is at rest in the laboratory system, $E_{2(\text{lab})} = M_pc^2$ and $\mathbf{p}_{2(\text{lab})} = 0$. Using

$$E_1{}^2{}_{(\text{lab})} - \mathbf{p}_1{}^2{}_{(\text{lab})}c^2 = M_p{}^2c^4,$$

we see that (43) reduces to

$$2E_{1(\text{lab})}M_pc^2 + 2M_p{}^2c^4 = E^2{}_{\text{tot(c.m.)}}, \tag{44}$$

where $E_{\text{tot(c.m.)}}$ denotes the sum $E_1 + E_2$ in the center-of-mass system. If we let $E_{\text{tot(lab)}}$ denote the total energy $E_1 + M_pc^2$ in the laboratory system, we have from (44):

$$2E_{\text{tot(lab)}}M_pc^2 = E^2{}_{\text{tot(c.m.)}}, \tag{45}$$

or

$$\boxed{\frac{E_{\text{tot(c.m.)}}}{E_{\text{tot(lab)}}} = \frac{2M_pc^2}{E_{\text{tot(c.m.)}}}.} \tag{46}$$

This is a measure of the "efficiency." To get a total energy of 20 Bev in the center-of-mass system, for $M_pc^2 \approx 1$ Bev, we need

$$E_{\text{tot(lab)}} = \frac{E_{\text{tot(c.m.)}}}{2M_pc^2} \approx \frac{400}{2} \approx 200 \text{ Bev}. \tag{47}$$

In this case about 20 Bev of the kinetic energy of the 200-Bev proton in the laboratory system is available to create new particles. Because of this low efficiency, consideration has been given to accelerator designs in which two beams of particles will collide from opposite directions, so that the laboratory frame will be the center-of-mass frame.

Antiproton Threshold. The energy of the Bevatron at Berkeley was designed to make possible the production of antiprotons (denoted by $\bar{p}$) by bombarding stationary protons with high-energy protons. The reaction may be written

$$p + p \rightarrow p + p + (p + \bar{p}); \tag{48}$$

that is, a proton-antiproton pair is produced. This conserves charge, because the antiproton carries charge $-e$. What is the threshold energy for the reaction?

The rest energy of a proton-antiproton pair is $2M_pc^2$, because the rest mass of the antiproton is equal to the rest mass of the proton. In the center-of-mass system the kinetic energy must therefore be at least $2M_pc^2$, which is M_pc^2 for each of the two initial protons. To this must be added the rest energy M_pc^2 for each of the initial protons, so that the minimum total energy in the center-of-mass system is

$$E_{\text{tot(c.m.)}} = 4M_pc^2. \tag{49}$$

In the laboratory system the corresponding energy is, from (46),

$$E_{\text{tot(lab)}} = \frac{E^2{}_{\text{tot(c.m.)}}}{2M_pc^2} \cong \frac{16}{2} M_pc^2, \tag{50}$$

of which $2M_pc^2$ is rest energy of the two protons and $6M_pc^2$ is kinetic energy. Thus the threshold energy is

$$6M_pc^2 = 6(0.938 \text{ Bev}) \cong 5.63 \text{ Bev}. \tag{51}$$

If the incident proton collides with a proton bound in a nucleus, the threshold energy is lower because the target proton is bound. Can you see why? The observed threshold energy for antiproton production is 4.4 Bev, which is 1.2 Bev lower than calculated for free target protons at rest. This threshold is the minimum kinetic energy required of the incident proton, as seen in the laboratory system, to make the reaction go.

Relativistic Rocket Equation. Suppose that, as viewed in the reference frame S′ in which the rocket is instantaneously at rest, the rocket experiences a constant acceleration a'. By the results of Chap. 11 we know that the increment of velocity $\Delta v'$ in S′ is related to the increment of velocity Δv in an inertial frame S by

$$\Delta v = \left(1 - \frac{v^2}{c^2}\right) \Delta v', \tag{52}$$

where v is the velocity of S′ relative to S. The frame S might be that of the earth. From (52), with the acceleration a' written for dv'/dt',

$$\frac{\Delta v}{1 - v^2/c^2} = \frac{dv'}{dt'} \Delta t' = a' \Delta t', \tag{53}$$

where $\Delta t'$ is the time interval as measured in the rocket frame S′. If the rocket starts at $t' = 0$ from rest with respect to S, we integrate the left- and right-hand sides of (53) to obtain

$$c^2 \int_0^v \frac{dv}{c^2 - v^2} = a' \int_0^{t'} dt'. \tag{54}$$

Using Dwight 728.4, we obtain

$$c \tanh^{-1}\left(\frac{v}{c}\right) = a't'; \qquad \frac{v}{c} = \tanh \frac{a't'}{c} \tag{55}$$

We find the total displacement of the rockets as viewed in S by a further integration:

$$x = \int_0^t v\, dt = \int_0^{t'} v \frac{dt'}{(1 - v^2/c^2)^{1/2}}. \tag{56}$$

Using (53),

$$dt' = \frac{1}{a'} \cdot \frac{c^2\,dv}{c^2 - v^2}, \tag{57}$$

so that (56) becomes

$$x = \frac{c^3}{a'} \int_0^v \frac{v}{(c^2 - v^2)^{3/2}}\,dv = \frac{c^3}{a'}\left[\frac{1}{(c^2 - v^2)^{1/2}} - \frac{1}{c}\right]. \tag{58}$$

Now by (11.18) and (11.20), we obtain from (55)

$$\cosh\frac{a't'}{c} = \frac{1}{(1 - v^2/c^2)^{1/2}}, \tag{59}$$

so that (58) becomes

$$x = \frac{c^2}{a'}\left(\cosh\frac{a't}{c} - 1\right). \tag{60}$$

From (60) we find that a man traveling on a rocket for 21 years, as measured in *his* reference frame S′, under a constant acceleration of 1g will cover a distance of 1.2×10^9 *light-years*, as measured in the reference frame S. How far would the man have traveled if $c \to \infty$, the limit of Galilean transformations?

Suggested reading

Lawrence Radiation Laboratory, ***Introduction to the detection of nuclear particles in a bubble chamber*** (Ealing Press, Cambridge, Mass., 1964). A wonderful collection of bubble-chamber photographs of particle tracks (complete with a stereoscopic viewer).

Film list

"The Ultimate Speed, an Exploration with High Energy Electrons," by William Bertozzi, M.I.T. (38 min. black and white). The experiment described in this film is an excellent demonstration of the acceleration of electrons by a longitudinal electric field, discussed in the first part of Chap. 13.

Problems

1. Calculate the gyroradius and gyrofrequency of a proton of total relativistic energy 10 Bev in a magnetic field of 10,000 gauss.

Ans. $\omega_c = 9 \times 10^6$ rad/sec.

2. What is the recoil energy in ergs and in electron volts of a nucleus of mass 10^{-23} gm after the emission of a γ ray of energy 1 Mev?

Ans. 1.4×10^{-10} erg; 90 ev.

3. An electron of energy 10 Bev collides with a proton at rest.

(*a*) What is the velocity of the center-of-mass system?

(*b*) What energy is available to produce new particles? (Express in units of M_pc^2.)

4. At high energies the cyclotron frequency depends on the speed of the particle which is being accelerated. In order to maintain synchronism between the cycling particle and the alternating electric field which accelerates it, the designer must require that the applied radio frequency or the magnetic field (or both) be modulated as the acceleration progresses. Show that $\omega \propto B/E$, where ω is the radio frequency, B is the magnetic field, and E is the total energy of the particle. [Use Eq. (26).]

5. In one stage of development the Berkeley 184-in. cyclotron operated at a fixed magnetic field of approximately 15,000 gauss.

(*a*) Calculate the nonrelativistic cyclotron frequency for protons in this field. *Ans.* 1.4×10^8 rad/sec.

(*b*) Calculate the frequency appropriate for a final kinetic energy of 300 Mev.

6. For the relativistic rocket problem we found in (59) that

$$\gamma = \frac{1}{(1 - v^2/c^2)^{1/2}} = \cosh \frac{a't'}{c}.$$

(*a*) Evaluate γ for $a' = 10^3$ cm/sec^2 and $t' = 10$ years $\cong 3 \times 10^8$ sec.

(*b*) If the rest mass of the rocket after this time is 1000 kg, what is the mass as observed in S? *Ans.* 1×10^7 kg.

(*c*) What is the kinetic energy in S? Could this amount of energy be produced feasibly? *Ans.* 1×10^{31} ergs.

7. (*a*) Show that a free electron moving in a vacuum at velocity **v** cannot emit a single light quantum. That is, show that such an emission process would violate the conservation laws.

(*b*) A hydrogen atom in an excited electronic state can emit a light quantum. Show that such a process can satisfy the conservation laws. What is the reason for the difference between parts (*a*) and (*b*)?

8. Calculate the momentum, total energy, and kinetic energy of a proton for which $\beta \equiv v/c = 0.99$ in the following cases:

(*a*) In the laboratory frame. *Ans.* 6.58 Bev/*c*; 6.63 Bev; 5.69 Bev.

(*b*) In a frame moving with the particle.

(*c*) In a frame stationary with respect to the center of mass of the proton and a stationary helium nucleus. ($M_{\text{He}} \cong 4M_p$.)

9. Find the radius of the orbit of a particle of charge e and of energy 10^{19} ev in a magnetic field of 10^{-6} gauss. (A magnetic field of 10^{-6} gauss is not unreasonable for a magnetic field in our galaxy.) Compare this with the diameter of our galaxy. (Particles causing "events" of this tremendous energy have been detected in cosmic rays; such particles give rise to what are called *extensive air showers* of electrons, positrons, γ rays, and mesons.)

10. (*a*) Calculate the radius of curvature in the path of a proton of kinetic energy 1 Bev in a transverse magnetic field of 20,000 gauss. *Ans.* 284 cm.

(*b*) What transverse electric field is required to produce about the same radius of curvature? Use the fact that the radius of curvature of a curve $y(x)$ is given by $\rho = [1 + (dy/dx)^2]^{3/2}/(d^2y/dx^2)$.

Ans. 1.75×10^4 statvolts/cm.

(*c*) Consider the magnitude of the electric field in part (*b*) and comment on the practicability of using electric fields for the deflection of relativistic particles.

11. Consider the nuclear reaction in which an incident proton of kinetic energy K_p strikes and splits a stationary deuteron according to

$$p + d \rightarrow p + p + n.$$

Near threshold the two protons and the neutron move in an unbound cluster with approximately the same velocity. Write down nonrelativistic expressions for momentum and energy and show that the threshold kinetic energy $K_p{}^0$ of the incident proton is

$$K_p{}^0 = \tfrac{3}{2}E_B,$$

where E_B (≈ 2 Mev) is the binding energy of the deuteron, with respect to a free neutron and proton.

12. When one calculates the threshold energy in a high-energy interaction, we have seen that it is convenient to examine the conditions in the center-of-mass system. Consider the reaction

$$\gamma + p \rightarrow p + \pi^0,$$

in which a photon γ is incident upon a stationary proton and produces a π^0 meson.

(*a*) Show that the threshold energy of the photon is

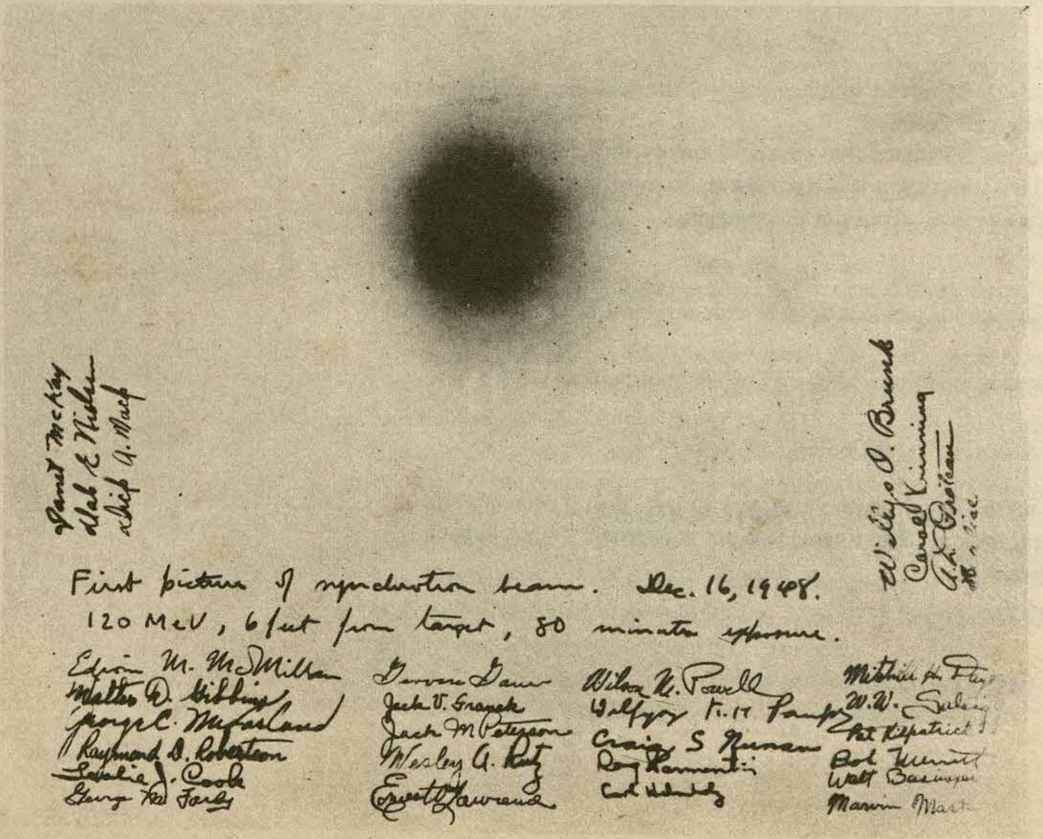
$$E_\gamma = M_\pi\left(1 + \frac{M_\pi}{2M_p}\right)c^2.$$

(*b*) Not all the initial photon energy is converted into rest mass energy of the neutral meson. What percent is not converted? What happens to the unconverted part?

Historical Note: Synchrotron

The synchrotron principle is utilized in all high-energy accelerators in the region over 1 Bev, except for electron linear accelerators such as that at Stanford. The synchrotron is a device to accelerate particles to high energies. It is essentially a cyclotron in which either the magnetic field or the applied radio frequency is varied during the acceleration, and in which the phase of the particles with respect to the rf electric field automatically adjusts itself to the proper value for acceleration. The idea of frequency or field modulation was not new at the time; what was new was the demonstration that particle orbits could be stable during the modulation. The synchrotron principle was discovered by V. Veksler in Moscow and independently a little later by E. M. McMillan in Berkeley. A full account of Veksler's work appeared in *Journal of Physics (USSR)* **9**, 153–158 (1945). McMillan's account appeared in the *Physical Review* **68**, 143 (1945). We reproduce McMillan's publication.

The first picture of the synchrotron beam.
(*Lawrence Radiation Laboratory photograph*)



First picture of synchrotron beam. Dec. 16, 1948.
120 MeV, 6 feet from target, 80 minute exposure.

Edwin M. McMillan
Walter D. Gibbins
George C. McFarland
Raymond A. Robertson
J. Cook
George W. Farly

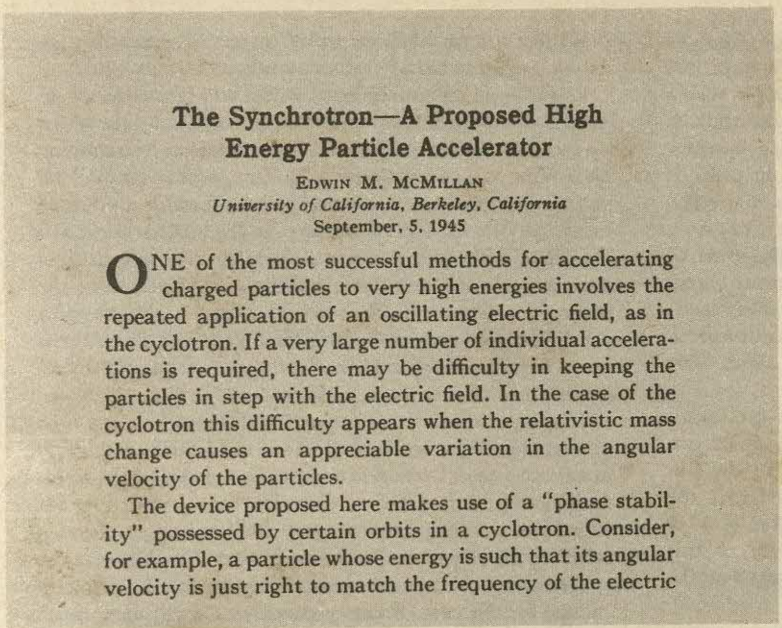
Jack V. Franck
Jack M Peterson
Wesley A. Reitz
Ernest Lawrence

Wilson M. Powell
Wolfgang K. H. Panofsky
Craig S Nunan
Roy Parmentier

W. W. Salsig
Pat Kilpatrick
Bob Merritt
Walt Baumgarten
Marvin Martin

Daniel McKay
Hal E. Nielsen
Rich A. Mack

Wilfred O. Brunk
Carol Kinning
A. K. Christian



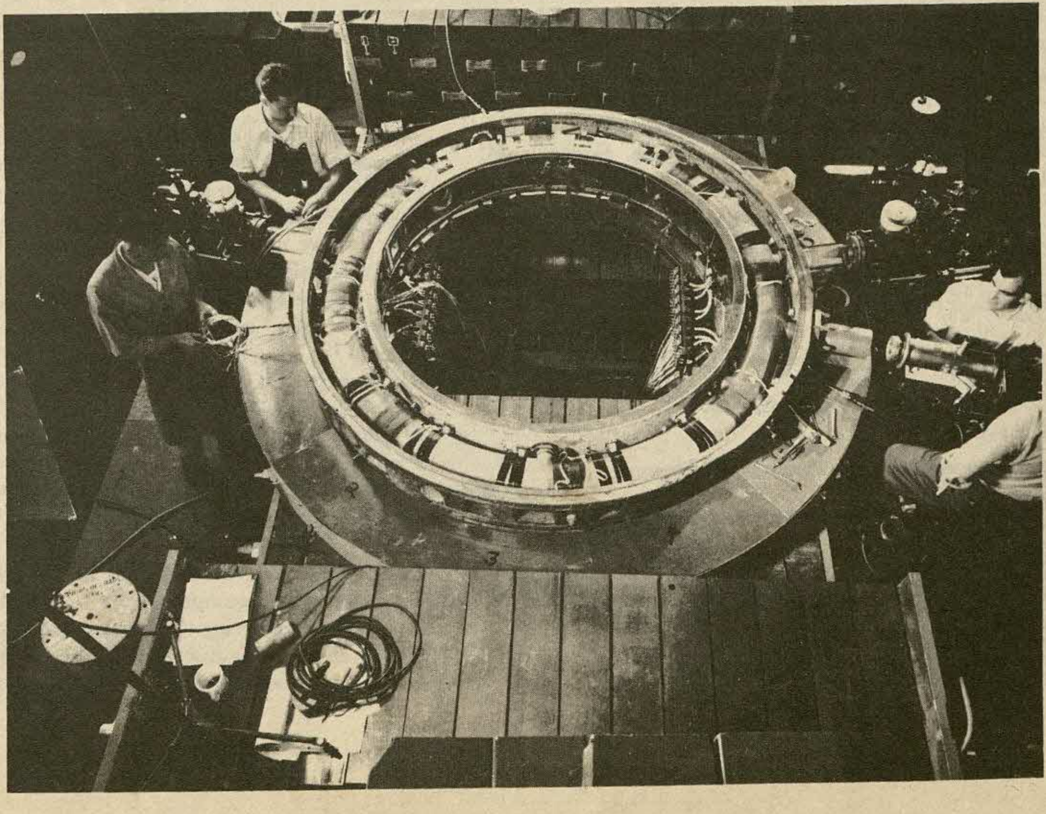
The Synchrotron—A Proposed High Energy Particle Accelerator

EDWIN M. MCMILLAN
University of California, Berkeley, California
September, 5, 1945

ONE of the most successful methods for accelerating charged particles to very high energies involves the repeated application of an oscillating electric field, as in the cyclotron. If a very large number of individual accelerations is required, there may be difficulty in keeping the particles in step with the electric field. In the case of the cyclotron this difficulty appears when the relativistic mass change causes an appreciable variation in the angular velocity of the particles.

The device proposed here makes use of a "phase stability" possessed by certain orbits in a cyclotron. Consider, for example, a particle whose energy is such that its angular velocity is just right to match the frequency of the electric

First electron synchrotron. (*Lawrence Radiation Laboratory photograph*)



field. This will be called the equilibrium energy. Suppose further that the particle crosses the accelerating gaps just as the electric field passes through zero, changing in such a sense that an earlier arrival of the particle would result in an acceleration. This orbit is obviously stationary. To show that it is stable, suppose that a displacement in phase is made such that the particle arrives at the gaps too early. It is then accelerated; the increase in energy causes a decrease in angular velocity, which makes the time of arrival tend to become later. A similar argument shows that a change of energy from the equilibrium value tends to correct itself. These displaced orbits will continue to oscillate, with both phase and energy varying about their equilibrium values.

In order to accelerate the particles it is now necessary to change the value of the equilibrium energy, which can be done by varying either the magnetic field or the frequency. While the equilibrium energy is changing, the phase of the motion will shift ahead just enough to provide the necessary accelerating force; the similarity of this behavior to that of a synchronous motor suggested the name of the device.

The equations describing the phase and energy variations have been derived by taking into account time variation of both magnetic field and frequency, acceleration by the "betatron effect" (rate of change of flux), variation of the latter with orbit radius during the oscillations, and energy losses by ionization or radiation. It was assumed that the period of the phase oscillations is long compared to the period of orbital motion. The charge was taken to be one electronic charge. Equation (1) defines the equilibrium energy; (2) gives the instantaneous energy in terms of the equilibrium value and the phase variation, and (3) is the "equation of motion" for the phase. Equation (4) determines the radius of the orbit.

$$E_0=(300cH)/(2\pi f), \tag{1}$$

$$E=E_0[1-(d\phi)/(d\theta)], \tag{2}$$

$$2\pi\frac{d}{d\theta}\left(E_0\frac{d\phi}{d\theta}\right)+V\sin\phi =\left[\frac{1}{f}\frac{dE_0}{dt}-\frac{300}{c}\frac{dF_0}{dt}+L\right]+\left[\frac{E_0}{f^2}\frac{df}{dt}\right]\frac{d\phi}{d\theta}, \tag{3}$$

$$R=(E^2-E_r^2)^{\frac{1}{2}}/300H. \tag{4}$$

The symbols are:

E = total energy of particle (kinetic plus rest energy),
E_0 = equilibrium value of E,
E_r = rest energy,
V = energy gain per turn from electric field, at most favorable phase for acceleration,
L = loss of energy per turn from ionization and radiation,
H = magnetic field at orbit,
F_0 = magnetic flux through equilibrium orbit,
ϕ = phase of particle (angular position with respect to gap when electric field $=0$),
θ = angular displacement of particle,
f = frequency of electric field,
c = light velocity,
R = radius of orbit.

(Energies are in electron volts, magnetic quantities in e.m.u., angles in radians, other quantities in c.g.s. units.)

Equation (3) is seen to be identical with the equation of motion of a pendulum of unrestricted amplitude, the terms on the right representing a constant torque and a damping force. The phase variation is, therefore, oscillatory so long as the amplitude is not too great, the allowable amplitude being $\pm\pi$ when the first bracket on the right is zero, and vanishing when that bracket is equal to V. According to the adiabatic theorem, the amplitude will diminish as the inverse fourth root of E_0, since E_0 occupies the role of a slowly varying mass in the first term of the equation; if the frequency is diminished, the last term on the right furnishes additional damping.

The application of the method will depend on the type of particles to be accelerated, since the initial energy will in any case be near the rest energy. In the case of electrons, E_0 will vary during the acceleration by a large factor. It is not practical at present to vary the frequency by such a large factor, so one would choose to vary H, which has the additional advantage that the orbit approaches a constant radius. In the case of heavy particles E_0 will vary much less; for example, in the acceleration of protons to 300 Mev it changes by 30 percent. Thus it may be practical to vary the frequency for heavy particle acceleration.

A possible design for a 300 Mev electron accelerator is outlined below:

peak $H=10{,}000$ gauss,
final radius of orbit $=100$ cm,
frequency $=48$ megacycles/sec.,
injection energy $=300$ kv,
initial radius of orbit $=78$ cm.

Since the radius expands 22 cm during the acceleration, the magnetic field needs to cover only a ring of this width, with of course some additional width to shape the field properly. The field should decrease with radius slightly in order to give radial and axial stability to the orbits. The total magnetic flux is about $\frac{1}{5}$ of what would be needed to satisfy the betatron flux condition for the same final energy.

The voltage needed on the accelerating electrodes depends on the rate of change of the magnetic field. If the magnet is excited at 60 cycles, the peak value of $(1/f)(dE_0/dt)$ is 2300 volts. (The betatron term containing dF_0/dt is about $\frac{1}{5}$ of this and will be neglected.) If we let $V=10{,}000$ volts, the greatest phase shift will be 13°. The number of turns per phase oscillation will vary from 22 to 440 during the acceleration. The relative variation of E_0 during one period of the phase oscillation will be 6.3 percent at the time of injection, and will then diminish. Therefore, the assumptions of slow variation during a period used in deriving the equations are valid. The energy loss by radiation is discussed in the letter following this, and is shown not to be serious in the above case.

The application to heavy particles will not be discussed in detail, but it seems probable that the best method will be the variation of frequency. Since this variation does not have to be extremely rapid, it could be accomplished by means of motor-driven mechanical turning devices.

The synchrotron offers the possibility of reaching energies in the billion-volt range with either electrons or heavy particles; in the former case, it will accomplish this end at a smaller cost in materials and power than the betatron; in the latter, it lacks the relativistic energy limit of the cyclotron.

Construction of a 300-Mev electron accelerator using the above principle at the Radiation Laboratory of the University of California at Berkeley is now being planned.

Chapter 14

Inertial and gravitational mass 427
Gravitational mass of photons 430
Gravitational red shift 433
Example: Deflection of photons by the sun 434
Equivalence 435

Problems 436

Historical Note: Newton's pendulums 437

Principle of Equivalence

Inertial and Gravitational Mass. We may determine the mass of a body by measuring the acceleration produced on it by a known force:

$$M_i = \frac{F}{a}. \tag{1}$$

The mass determined in this way is known as the *inertial mass* and is denoted by M_i. We may also determine the mass by measuring the gravitational force exerted on it by another body, such as the earth:

$$\begin{gathered} \frac{GM_gM_E}{r^2} = F; \\ M_g = \frac{Fr^2}{GM_E}. \end{gathered} \tag{2}$$

The mass determined in this way is known as the *gravitational mass* and is denoted by M_g. In (2) the mass of the earth is M_E.

It is a remarkable fact that the inertial mass of all bodies is, within experimental accuracy, proportional to the gravitational mass. (By a suitable choice of G, the inertial mass may be brought to equal the gravitational mass.) The simplest experiment to check this is to see whether all bodies fall with the same acceleration. For one falling body near the surface of the earth, we have

$$M_i(1)a(1) = \frac{GM_EM_g(1)}{R_E{}^2}; \tag{3}$$

for a second falling body,

$$M_i(2)a(2) = \frac{GM_E M_g(2)}{R_E^2}. \tag{4}$$

On dividing Eq. (3) by Eq. (4) we have

$$\begin{aligned} \frac{M_i(1)a(1)}{M_i(2)a(2)} &= \frac{M_g(1)}{M_g(2)}; \\ \frac{M_i(1)}{M_g(1)} &= \frac{M_i(2)}{M_g(2)} \cdot \frac{a(2)}{a(1)}. \end{aligned} \tag{5}$$

But falling bodies in a vacuum are always observed to fall at the same rate, so that $a(2) = a(1)$ within the experimental accuracy, and therefore we have for the ratio of inertial to gravitational mass

$$\frac{M_i(1)}{M_g(1)} = \frac{M_i(2)}{M_g(2)}. \tag{6}$$

As long as this ratio of masses is constant, we can always make the value of the ratios in (6) equal to unity by adjusting appropriately the value of G. The experimental task is to determine whether variations of the ratio M_i/M_g exist for different particles, materials, and objects.

The classical determinations were carried out by Newton, using the pendulum method of Prob. 1. Other famous determinations include those by R. Eötvös, which he started about 1890 and continued for about 25 years. His ingenious method can be understood by considering a pendulum suspended at the surface of the earth and at a latitude of 45°. The pendulum is acted on by a gravitational force $M_g g$ directed toward the center of the earth. It is acted on also by a centrifugal force $M_i\omega^2 R_E/\sqrt{2}$, where the factor $1/\sqrt{2}$ enters as the cosine of 45°; note that $R_E/\sqrt{2}$ is the perpendicular distance at this latitude from the pendulum to the axis of rotation of the earth. The centrifugal force is directed normal to the axis of rotation. The resultant of the two forces makes an angle

$$\theta \cong \frac{M_i\omega^2 R_E/2}{M_g g - \frac{1}{2}M_i\omega^2 R_E} \cong \frac{M_i\omega^2 R_E}{2M_g g} \tag{7}$$

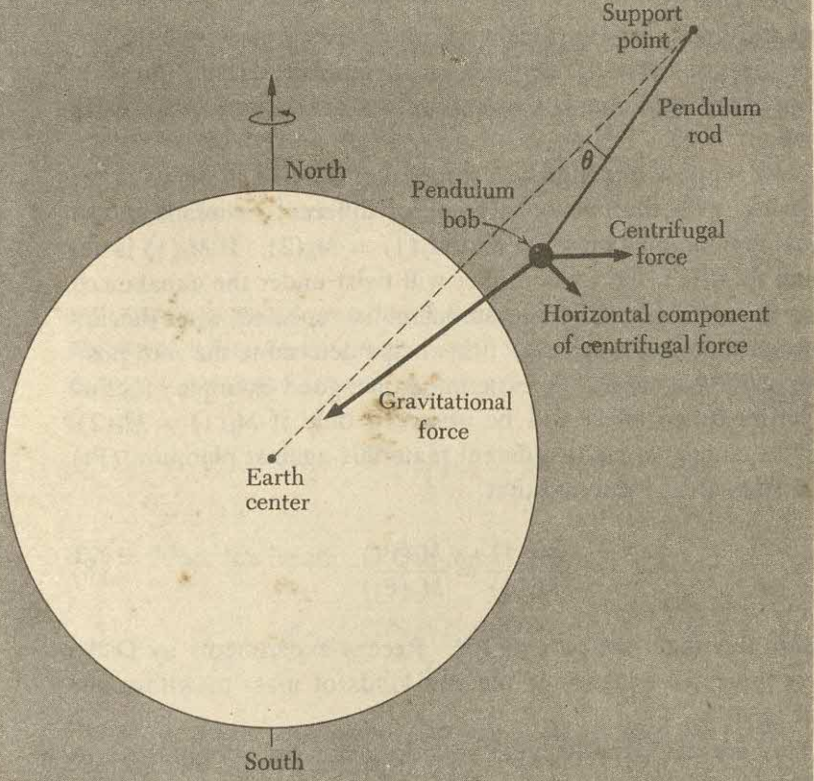
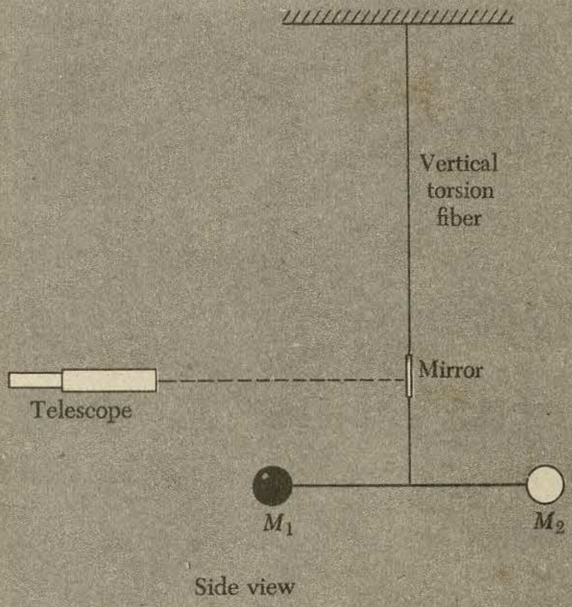
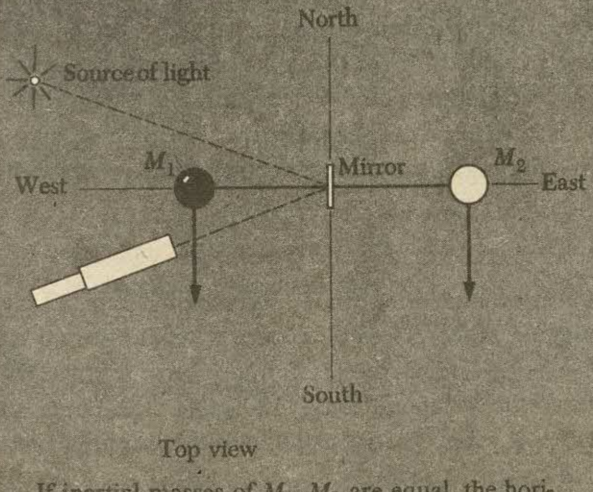


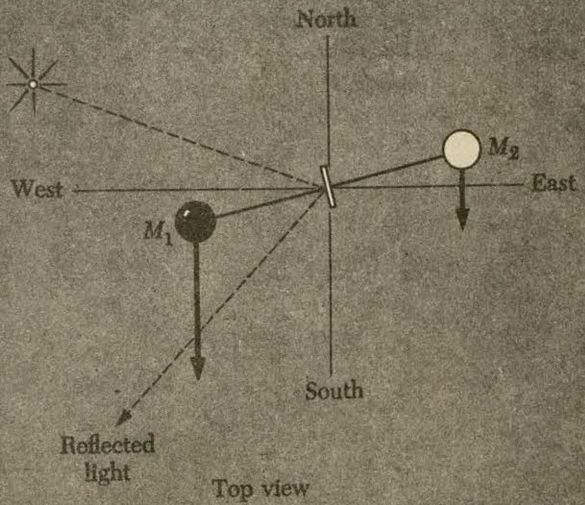
Illustration of how a pendulum is deflected from vertical by small angle θ because of centrifugal force arising from earth's rotation.



Side view of an apparatus similar to that employed by Eötvös to determine the ratio of inertial to gravitational mass. M_1 and M_2 are two dissimilar objects of the same gravitational mass.



If inertial masses of M_1, M_2 are equal, the horizontal components of centrifugal force (two arrows) are equal, and there is no net torsion on fiber.



If inertial mass of M_1 is greater than that of M_2, there will be a twist in the fiber, and the mirror will rotate.

with the direction of the local vertical. Here we have used the fact that the ratio $M_i\omega^2R_E/M_gg$ is a small number. Using the data given at the beginning of Chap. 3, we see that the ratio has a value of about 0.003.

Now suppose that a torsion suspension is made as shown in the drawing, with the two bobs made of different materials but of equal gravitational mass, so that $M_g(1) = M_g(2)$. If $M_i(1)$ is not equal to $M_i(2)$, the torsion fiber will twist under the unbalanced centrifugal force. The measurement is repeated with the apparatus turned through 180°; this helps determine the zero position of the balance. The experiment is a good example of a null experiment: an effect will be observed only if $M_i(1) \neq M_i(2)$. Eötvös compared eight different materials against platinum (Pt) as a standard. He found that

$$\frac{M_i(1)}{M_g(1)} = \frac{M_i(\text{Pt})}{M_g(\text{Pt})}, \tag{8}$$

within less than one part in 10^8. Recent experiments by Dicke have indicated equality of the two kinds of mass to within one part in 10^{10}.

The present experimental situation can be summarized as follows:

If we denote the ratio M_g/M_i by Q, then

(*a*) The value of Q for an electron plus a proton equals the value of Q for a neutron up to one part in 10^7. (This comparison follows directly from a comparison of light and heavy elements in the periodic table; heavy elements have a higher proportion of neutrons than light elements have.)

(*b*) The value of Q for that part of the nuclear mass associated with nuclear binding equals one within one part in 10^5.

(*c*) The value of Q for that part of the atomic mass associated with the binding of the orbital electrons equals one within one part in 200.

Gravitational Mass of Photons. We saw in Chap. 12 that a photon of energy $h\nu$, where ν is the frequency, must have an inertial mass equal to $h\nu/c^2$. Does the photon also have a gravitational mass? Experimental evidence strongly indicates that it does, and that the gravitational mass is equal in value to the inertial mass. (The *rest* mass, of course, is zero.)

Consider a photon which, when at a height L above the surface of the earth, has frequency ν and energy $h\nu$. The energy of the photon is increased by MgL on falling through the distance L. The energy of the photon becomes $h\nu'$ where

$$h\nu' \cong h\nu + \frac{h\nu}{c^2} gL, \tag{9}$$

assuming a constant mass $h\nu/c^2$ for the photon during the fall (the argument being that ν' is not much different from ν). The frequency ν' measured for the photon *after* the fall is then, from (9),

$$\nu' \cong \nu\left(1 + \frac{gL}{c^2}\right). \tag{10}$$

If $L = 20$ m, the fractional frequency shift is

$$\frac{\Delta\nu}{\nu} = \frac{gL}{c^2} \approx \frac{(10^3)(2 \times 10^3)}{(3 \times 10^{10})^2} \approx 2 \times 10^{-15}. \tag{11}$$

This fantastically small effect has actually been observed by Pound and Rebka† using a γ-ray source. They find, with $\Delta\nu = \nu' - \nu$,

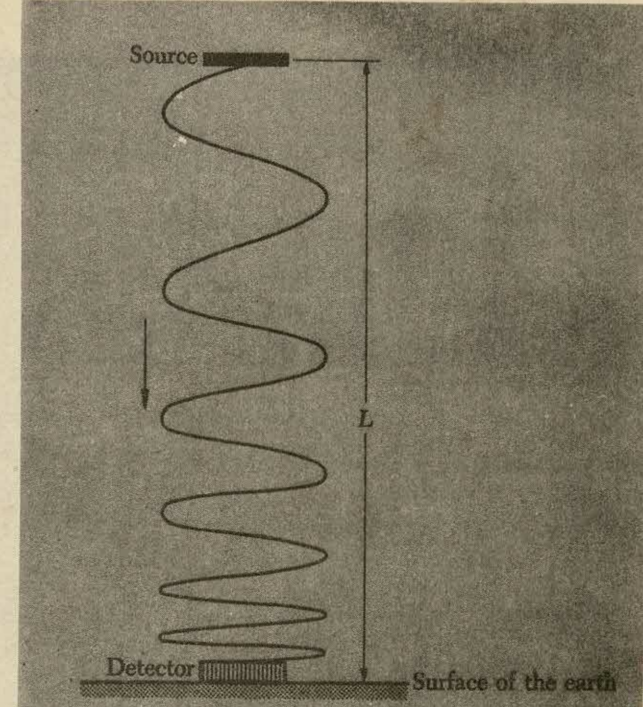
$$\frac{(\Delta\nu)_{\text{exp}}}{(\Delta\nu)_{\text{calc}}} = 1.05 \pm 0.10, \tag{12}$$

where the calculated value is obtained from (10).

A photon with frequency ν emitted at an infinite distance from the earth will have the frequency ν' on reaching the surface of the earth, where, by an obvious generalization of (9) and (10),

$$\nu' \cong \nu\left(1 + \frac{GM_E}{R_E c^2}\right). \tag{13}$$

Note that the frequency shift involves the ratio of the "gravitational length" GM_E/c^2 of the earth (as defined in Chap. 9) to the radius R_E of the earth. This ratio has the value 6×10^{-10}. The larger effect here is of the same kind as that considered in (11), but now the light source is much farther from the earth.

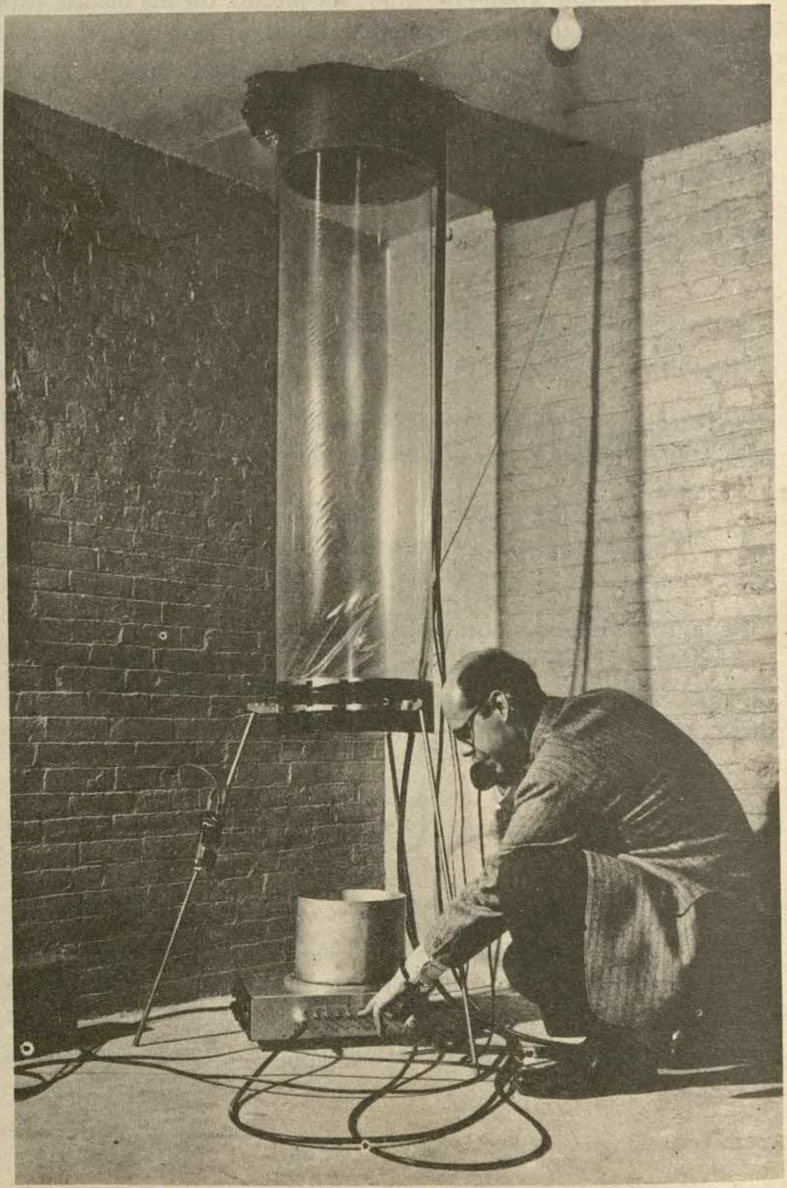


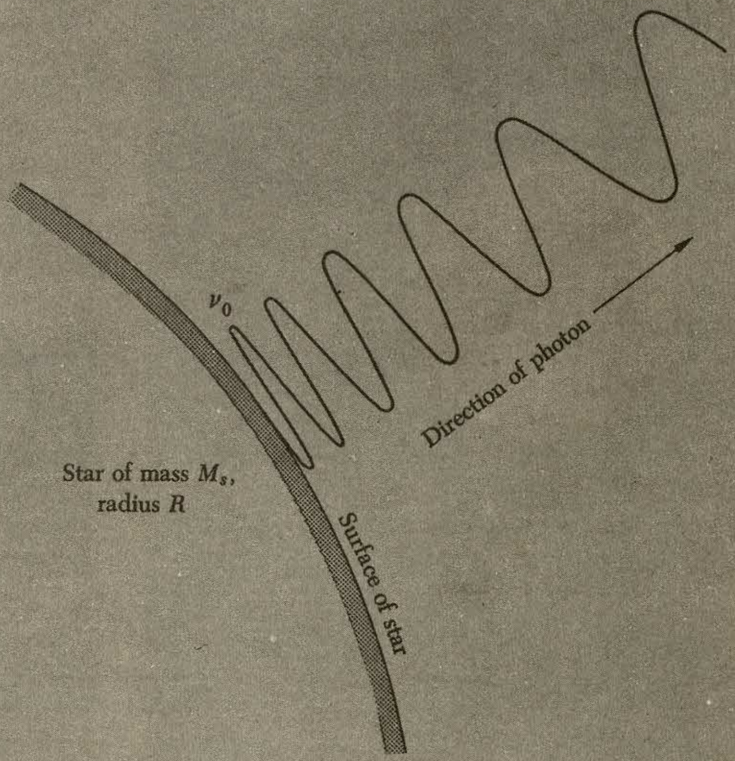
Schematic picture of "gravitational red-shift" experiment. A photon emitted at the source in a direction toward the earth's center loses "potential energy" $\Delta U = (h\nu/c^2)gL$ and gains an equal amount of "kinetic energy" in falling a distance L. The photon frequency at the detector is $\nu' = \nu(1 + gL/c^2)$; the photon frequency at the source is ν.

---

† R. V. Pound and G. A. Rebka, Jr., *Phys. Rev. Letters* **4**, 337 (1960).

Lower end of the Pound falling-photon experiment at Harvard, showing G. A. Rebka, Jr., adjusting photomultipliers on instruction from the control center. In a later version of the experiment means are provided for controlling the temperature of the source and absorber. The whole gravitational shift measured is only about 1/500 of the line width. To measure with accuracy such a small shift requires the aid of a few tricks. (*Courtesy of R. V. Pound*)





A photon escaping to infinity from the surface of a star gains "potential energy" and loses an equal amount of "kinetic energy." If the photon frequency at the surface is ν, the photon frequency at ∞ is $\nu' = \nu(1 - GM_s/R_sc^2)$.

Gravitational Red Shift. A photon of frequency ν which leaves a star and escapes to infinity will be observed at infinity with a frequency

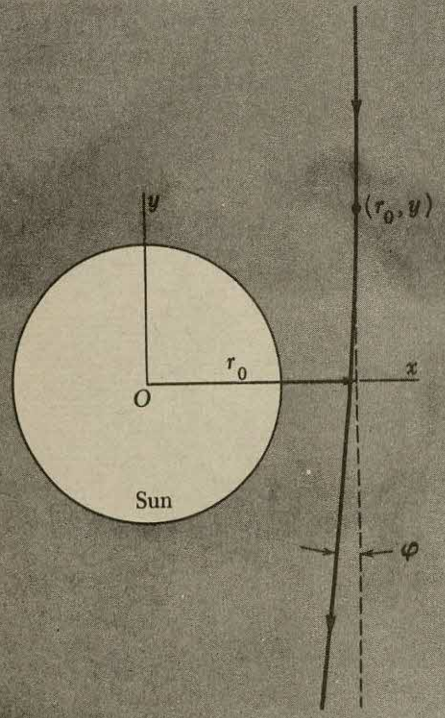
$$\nu' \cong \nu\left(1 - \frac{GM_s}{R_sc^2}\right), \tag{14}$$

where M_s is the mass and R_s the radius of the star. This follows by a modification of (13); the photon now has lost energy in escaping from the gravitational field of the star. A photon in the blue region of the visible spectrum will be shifted in frequency toward the red end of the spectrum: for this reason the effect is known as the *gravitational red shift.* It must not be confused with the recessional red shift of distant stars believed to arise from their apparent radial motion away from the earth, as discussed in Chap. 10.

White dwarf stars have large values of M_s/R_s and thus have relatively large values of the gravitational red shift. For Sirius B the calculated fractional shift is

$$\frac{\Delta\nu}{\nu} \cong -5.9 \times 10^{-5}, \tag{15}$$

and the observed value is -6.6×10^{-5}. The discrepancy is within the uncertainties in M_s and R_s.



Deflection of a photon by the gravitational field of the sun.

Example. Deflection of photons by the sun. What is the angular deflection of a light beam or photon which passes by the sun at its edge?

This problem involves a photon moving with the velocity of light in a gravitational field. We do not get the correct answer without doing a careful calculation using special relativity, but we can get the order of magnitude of the correct answer by a naïve calculation.

Suppose that the photon has a mass M_L; it will turn out that M_L drops out of the calculation of the deflection and thus we do not have to know what it is. Let the light beam pass the sun at a distance of closest approach r_0, as measured from the center of the sun. We suppose that the deflection will turn out to be very small, so that r_0 is essentially the same as if the light beam were not deflected. The transverse force F_x on the photon at the position (r_0,y) is

$$F_x = -GM_sM_L\frac{r_0}{(r_0^2 + y^2)^{3/2}}, \tag{16}$$

where y is measured from the point P as in the figure.

The final value of the transverse velocity component v_x of the photon has the value given by

$$M_Lv_x = \int F_x\,dt = \int F_x\frac{dy}{v_y} \cong \frac{1}{c}\int F_x\,dy, \tag{17}$$

so that

$$v_x \cong -\frac{2GM_sr_0}{c}\int_0^\infty \frac{dy}{(r_0^2 + y^2)^{3/2}} \cong -\frac{2GM_s}{cr_0}. \tag{18}$$

It follows that when r_0 is equal to the radius R_s of the sun, the angular deflection is

$$\varphi \cong \frac{|v_x|}{c} \cong \frac{2GM_s}{R_s c^2} \text{ radians.} \tag{19}$$

On doing the calculation, we find $\varphi = 0.87''$. A more careful study† using special relativity and the equivalence principle predicts twice the value given by our argument, or $1.75''$. This value has been confirmed by observation with an accuracy of perhaps 20 percent. (One still hears grumbling about the data.)

When we solve a collision problem by calculating the force on one particle as if its trajectory were a straight line, we are making what is called an *impulse approximation.* The connection between $\int F_x\,dt$ and the x component of the change in momentum is discussed in Chap. 5. The impulse approximation is often very useful, provided that the actual trajectory does not depart greatly from the straight line the particle would follow if there were no interaction.

Equivalence. The experimental result that no difference has ever been detected between the inertial mass and the gravitational mass of a body suggests that gravitation in a sense may be equivalent to acceleration. Consider an observer in an elevator which is freely falling with the acceleration g.

The equivalence principle states that to an observer in a freely falling elevator the laws of physics are the same as in the inertial frames of special relativity (at least in the immediate neighborhood of the center of the elevator). *The effects due to the accelerated motion and to the gravitational forces exactly cancel.* An observer sitting in an enclosed elevator cannot, if he observes apparent gravitational forces, tell what portion of these correspond to acceleration and what portion to actual gravitational forces. He will detect no forces at all *unless* other forces (i.e., other than gravitational forces) act on the elevator. In particular, the postulated principle of equivalence requires that the ratio of the inertial and gravitational masses be $M_i/M_g \equiv 1$. The "weightlessness" of a man in orbit in a satellite is a consequence of the equivalence principle.

---

† See, for example, L. I. Schiff, *Am. J. Phys.* **28**, 340 (1961).

Pursuit of the mathematical consequences of the principle of equivalence leads us to the general theory of relativity; for further discussion, consult the references suggested at the end of Chap. 11. The classical experimental tests of the general theory of relativity are discussed carefully in the first chapter of L. Witten, *Gravitation: An introduction to current research* (John Wiley and Sons, New York, 1962).

Problems

1. Show that the frequency of a pendulum of length L is given by

$$\nu = \frac{\omega}{2\pi} = \frac{1}{2\pi}\left(\frac{M_g}{M_i}\cdot\frac{g}{L}\right)^{1/2},$$

where M_g, M_i are the gravitational and inertial masses. (Bessel in the early days made careful pendulum observations and showed that M_g was equal to M_i within one part in 6×10^4.)

2. Find an expression for the gravitational red shift in which you do not use the assumption that $\Delta\nu/\nu \ll 1$. (Neglect any effects associated with the curvature of space.) Start with $h\,\Delta\nu = -(h\nu/c^2)(M_sG/r^2)\,\Delta r$, and integrate over dr from R_s to infinity, and over $d\nu$ from ν to ν'.

Ans. $\nu' = \nu e^{-GM_s/R_sc^2}$.

3. Estimate the gravitational red shift for light leaving the center of our galaxy, as observed far outside the galaxy. (Treat the distribution of mass as uniform within a sphere of radius 10,000 parsecs. The mass of the galaxy is $\sim 8 \times 10^{44}$ gm.) *Ans.* $\Delta\nu/\nu = -3 \times 10^{-6}$.

4. In 1962 an intense extraterrestrial source of radio radiation was optically identified as a starlike object with an angular radius of approximately $\frac{1}{2}$ sec of arc. It was at first thought to be a star in our galaxy giving off radio waves, but subsequently its spectrum was obtained and its spectral lines were found to be very considerably red-shifted. For instance, an atomic oxygen line with wavelength λ normally 3.727×10^{-5} cm, was identified at $\lambda = 5.097 \times 10^{-5}$ cm. One explanation took it to be an exceedingly massive star with a spectrum *gravitationally red-shifted.* If this hypothetical radio star is in our galaxy, its distance must be less than 10^{22} cm from the earth.

(*a*) Calculate from the angular diameter and the red shift, the mass and mean density of the star under this hypothesis, assuming the distance to be 10^{22} cm. Is this a reasonable explanation of this object?

(*b*) An alternative suggestion was that it might be a peculiar "radio galaxy," with its red shift following the usual recessional red-shift relation given in Chap. 10. Calculate its distance on this second hypothesis.

(*c*) Does the radio-galaxy hypothesis conform with this expectation?

Answers: (*a*) Mass is 8.7×10^{43} gm, mean density is 1.6×10^{-6} gm/cm^3. This does not seem reasonable, as the mass is about 0.1 of the total mass of our galaxy.

(*b*) 3×10^9 light-years (3×10^{27} cm).

(*c*) Yes; it has a radius of about 10^{22} cm. This is in the usual range of galactic radii.

[For further details see J. L. Greenstein, "Quasi-stellar radio sources," *Sci. American* **209,** 54 (1963).]

Historical Note. Newton's Pendulums

We quote from Newton's account in the *Principia* of his experiments with pendulums to investigate possible variations in the ratio of gravitational and inertial masses.

"But it has been long ago observed by others, that (allowance being made for the small resistance of the air) all bodies descended through equal spaces in equal times; and, by the help of pendulums, that equality of times may be distinguished to great exactness.

"I tried the thing in gold, silver, lead, glass, sand, common salt, wood, water, and wheat. I provided two equal wooden boxes. I filled the one with wood, and suspended an equal weight of gold (as exactly as I could) in the centre of oscillation of the other. The boxes, hung by equal threads of 11 feet, made a couple of pendulums perfectly equal in weight and figure, and equally exposed to the resistance of the air: and, placing the one by the other, I observed them to play together forwards and backwards for a long while, with equal vibrations. And therefore (by Cor. I and VI, Prop. XXIV, Book II) the quantity of matter in the gold was to the quantity of matter in the wood as the action of the motive force upon all the gold to the action of the same upon all the wood; that is, as the weight of the one to the weight of the other.

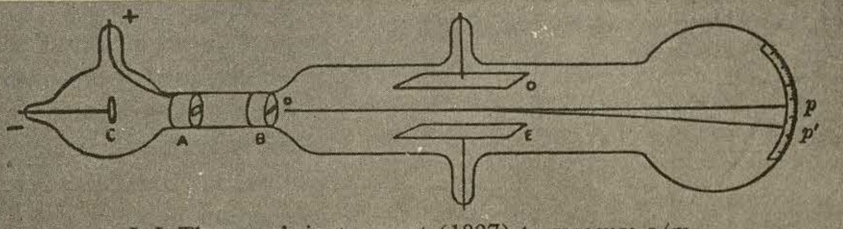
"And by these experiments, in bodies of the same weight, could have discovered a difference of matter less than the thousandth part of the whole."

Chapter 15

Stable and unstable particles 439
Masses 448
Charge 451
Lifetime 454
Other characteristics 456
Four basic forces of nature 457
Further reading 458
Problems 458
Historical Note 1: Rutherford scattering and the nuclear atom 459
Historical Note 2: Bubble chambers 468

The Particles of Modern Physics

Stable and Unstable Particles. All known matter is composed of particles. Perhaps the single most important discovery that physicists have made about the universe is the granularity of matter. This is the key to the behavior and structure of gases, liquids, and solids, to chemical reactions, to the kinds of models that can explain not only subatomic and atomic physics but macroscopic physics as well. As early as 1756 Franklin was conscious of the granularity of matter bearing electric charges. Upon contemplating the phenomenon of electrostatic induction, he said with amazing insight, "The electric matter consists of particles extremely subtle since it can permeate common matter, even the densest, with such freedom and ease as not to receive any appreciable resistance." In 1897 J. J. Thomson proved that cathode rays could be deflected by electrostatic and magnetic fields. He then calculated the mass of the particles (electrons) constituting these rays to be of the order of 10^{-3} of the mass of a hydrogen atom. This work, together with the simultaneous discovery of the Zeeman effect (splitting of spectral lines by a magnetic field), marks the wide acceptance of the electrical constitution of matter.

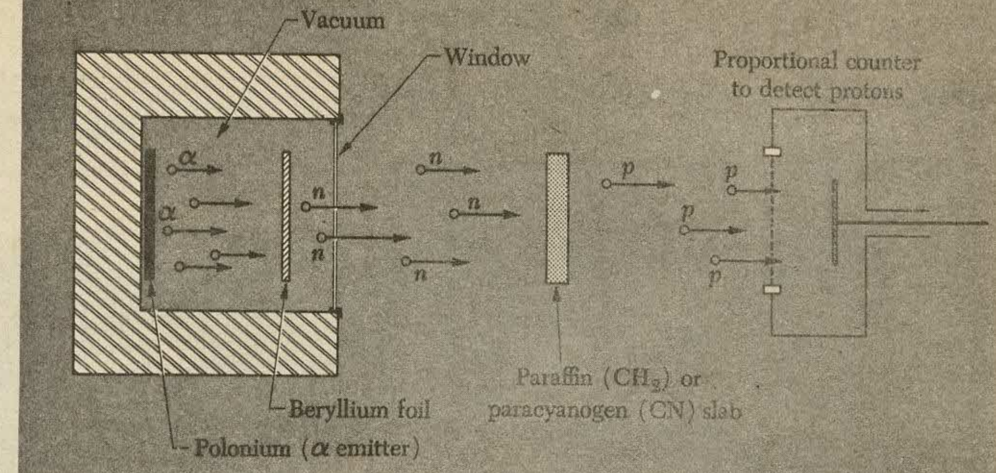


J. J. Thomson's instrument (1897) to measure e/m.

The Cavendish Laboratory in Cambridge University. Here the electron was discovered by J. J. Thomson, and the first artificial transmutation of a nucleus was observed by Rutherford. (*British Information Services photograph*)

The discovery of the nucleus by Rutherford followed in 1911–1913. Rutherford showed that the atoms consist of a positively charged nucleus surrounded by a system of electrons kept together by attractive Coulomb forces from the nucleus. It was also shown that the total negative charge of the electrons on an atom is equal to the positive charge of the nucleus. The nucleus is the seat of the greatest part of the mass of the atom; the linear dimensions of the nucleus are exceedingly small compared with the linear dimensions of the whole atom. As Bohr said in 1913, "Great interest is to be attributed to this atom-model."

After the development in 1913 of the Bohr-Rutherford picture of the atom, the known elementary particles were the electron, the photon, and some 95 different nuclei. (For practical purposes an elementary particle is any particle which cannot be easily described as a combination of other particles.) With the discovery of the neutron (a neutral particle of mass slightly greater than that


Vacuum
Window
Proportional counter to detect protons
α
α
n
n
n
n
n
p
p
p
p
Beryllium foil
Polonium (α emitter)
Paraffin (CH_2) or paracyanogen (CN) slab

Schematic diagram of apparatus used by Chadwick to observe neutrons (1932). In this experiment α rays from polonium strike a beryllium foil. A nuclear reaction

$$ {}_2\alpha^4 + {}_4\mathrm{Be}^9 \rightarrow {}_6\mathrm{C}^{13} \rightarrow {}_6\mathrm{C}^{12} + n $$

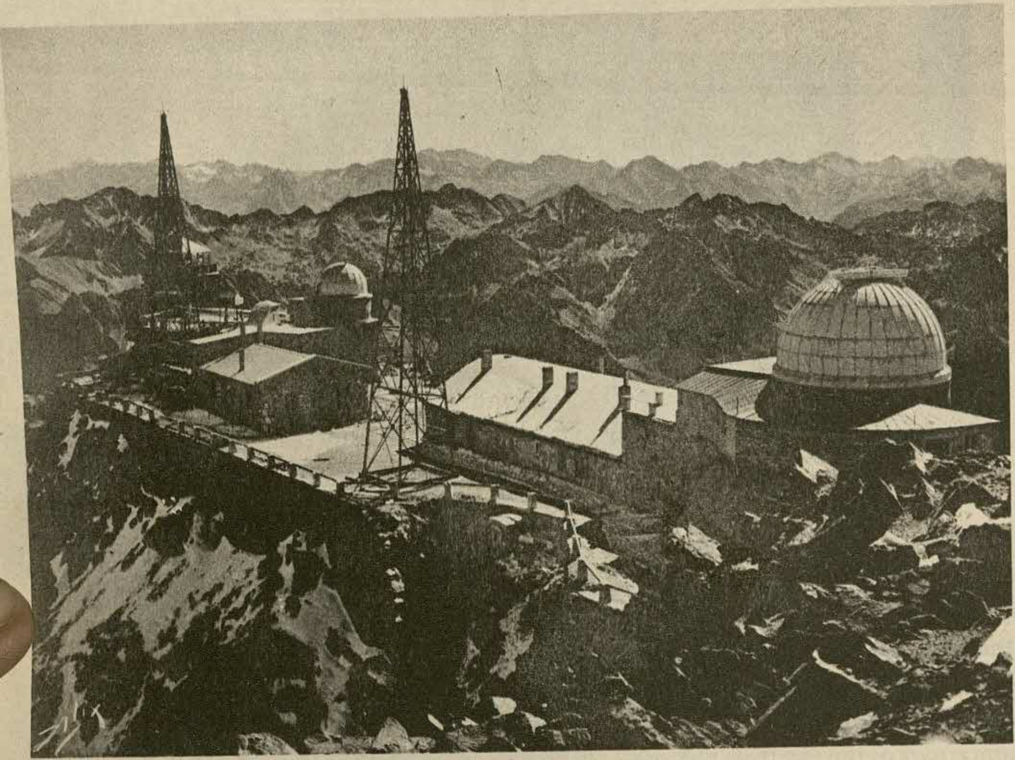
liberates neutrons which in turn strike a paraffin slab. The neutrons give up their energy to protons in elastic collisions; the protons are detected by means of a proportional counter. It was also possible to detect nitrogen nuclei with the proportional counter, when CN was used instead of CH_2. From this experiment, the mass of the neutron was determined to be approximately the mass of the proton.

of the proton) in 1932, the nuclei themselves were understood to be combinations of neutrons and protons; at that time four elementary particles were known: the neutron (n), the proton (p), the electron (e) and the photon (γ). Since that time the number of elementary particles known has grown continuously. The positron (e^+), a positively charged counterpart of the electron, was discovered in 1932. The existence of neutrinos (ν), which are massless neutral particles with very weak interactions with all other particles, was established reasonably well during the 1930s. Positively and negatively charged particles or muons (μ^+ and μ^-), about 208 times more massive than the electron, were discovered in 1936. Then the pace of discovery quickened: another pair of charged particles (π^+ and π^-, called pions or pi mesons†) in 1947; a neutral pion counterpart (π^0) in 1950; the negatively charged counterpart to the proton (antiproton, $\bar{p}$) in 1955; the antineutron ($\bar{n}$); a new family of charged and neutral mesons (K^+, K^0, K^-) and heavier particles (Λ^0, Σ^+, Σ^-, Σ^0, Ξ^0, Ξ^-); antiparticles to all of these; tens of very short-lived ($<10^{-20}$ sec) mesons and heavy particles; and even another family of neutrinos.

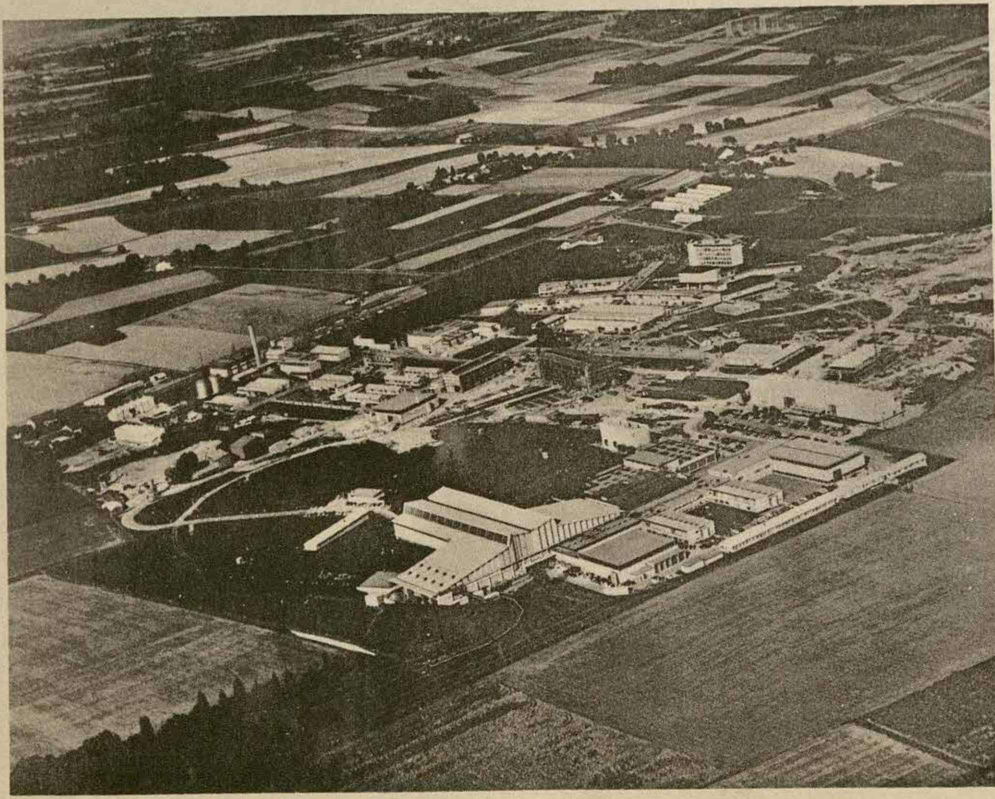
Can all these really be elementary particles? Are some or perhaps all of them interpretable as combinations of other particles, as bound states of two or more other particles? Could these particles be excited states of one particle, just as the hydrogen atom has many excited states? It is very likely that the great number of elementary particles is a reflection of our present ignorance of the answers to these questions.

---

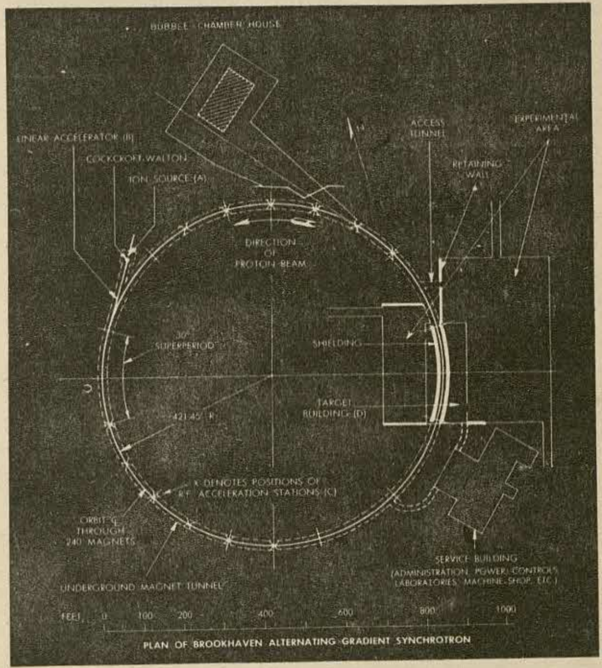
† The word *meson* means literally "intermediate," referring to the fact that its mass lies between that of the electron and the proton.



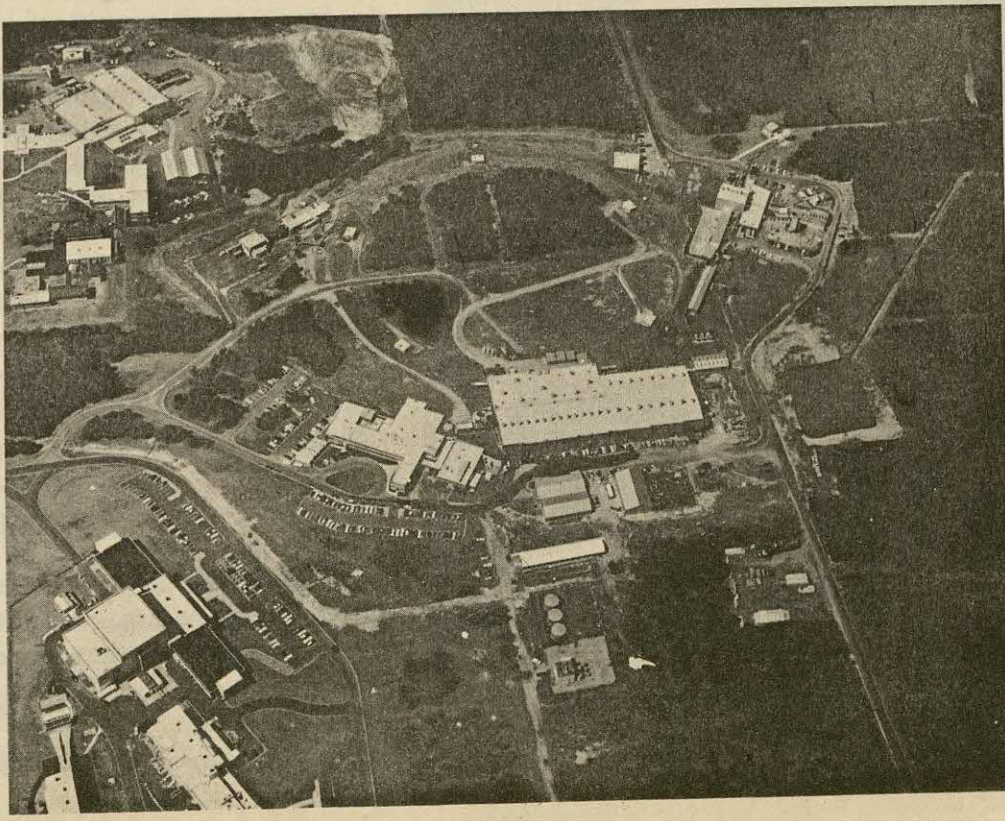
Pic du Midi Observatory in the French Pyrenees. Altitude 9,300 feet. Dome on left contains coronograph. The cosmic ray lab is just behind the dome. The skyline is the international boundary with Spain.



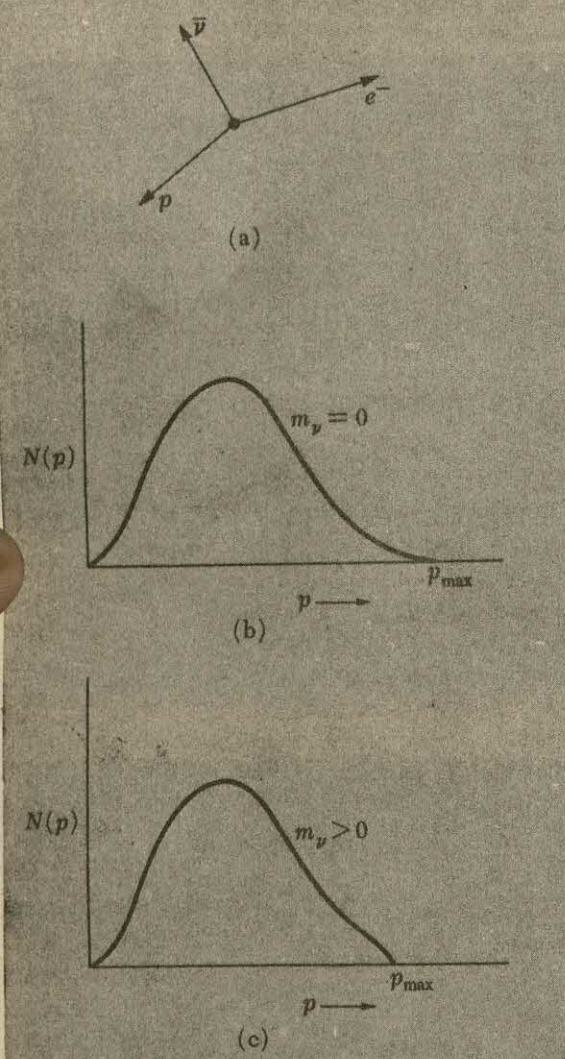
Aerial view of the CERN nuclear laboratories in Geneva. At the left can be seen the 28-Bev alternating gradient proton synchrotron, 200 m in diameter, embedded in earth, in which protons are accelerated up to 0.9994 of the speed of light. This machine is the joint realization of the fourteen European member states of CERN. (*CERN*)


BUBBLE CHAMBER HOUSE
LINEAR ACCELERATOR (B)
COCKCROFT-WALTON
ION SOURCE (A)
ACCESS TUNNEL
EXPERIMENTAL AREA
RETAINING WALL
DIRECTION OF PROTON BEAM
30° SUPERPERIOD
SHIELDING
TARGET BUILDING (D)
421.45' R
X DENOTES POSITIONS OF R.F. ACCELERATION STATIONS (C)
ORBIT ℄ THROUGH 240 MAGNETS
SERVICE BUILDING (ADMINISTRATION, POWER, CONTROLS, LABORATORIES, MACHINE-SHOP, ETC.)
UNDERGROUND MAGNET TUNNEL
FEET 0 100 200 400 600 800 1000
PLAN OF BROOKHAVEN ALTERNATING GRADIENT SYNCHROTRON

Plan of the 33-Bev proton synchrotron at Brookhaven National Laboratory. The high-energy particle accelerators at Brookhaven, Geneva, Dubna, and Berkeley are among the most impressive scientific tools in the world. Other great accelerators are under construction.



Aerial view of the Brookhaven alternating gradient synchrotron and its ancillary buildings: on the right side of the magnet ring is the 80-in. liquid hydrogen bubble-chamber building; *upper left,* the Cosmotron building; *lower left,* the graphite research-reactor building. (*Brookhaven National Laboratory*)



The free neutron decays to a proton and an electron (which we can observe) plus an antineutrino, which is practically impossible to observe (*a*). We know the antineutrino is there because the vector sum of the proton and electron momenta is not in general equal to the momentum of the original neutron. The number $N(p)$ of electrons per unit momentum range is plotted vs electron momentum (*b*). If the neutrino mass were greater than zero, the curve in (*c*) would be obtained; in fact, such distributions as that shown in (*c*) are never observed.

Only two of the charged particles, the electron and the proton, are stable—at least in our part of the universe. In a universe of antiparticles the positron and the antiproton would be stable. All the other charged particles either decay rapidly and spontaneously, generally in much less than 10^{-6} sec, or are almost instantly annihilated (positrons, antiprotons) when they come in contact with normal matter. Among the neutral (uncharged) particles only the photon and the neutrinos do not spontaneously decay. Free neutrons decay according to $n = p + e + \bar{\nu}$, with a mean life of 15 min, but neutrons bound in a nucleus do not necessarily decay. Bound neutrons and protons can form stable nuclei.

Thus, despite the great number of elementary particles, only a few play an obvious role in the structure of normal matter. Neutrons and protons are bound together to form charged nuclei. Around a nucleus there moves a cloud of electrons, and together they form an atom. Atoms combine into molecules. Vast collections of molecules form macroscopic forms of matter: gases, liquids, crystals. . . . Accelerated electrons emit or absorb photons. The study of transitions between atomic stationary states is spectroscopy, which together with electron and nuclear resonance has been the main experimental tool for exploring the structure of atoms and molecules.

Except for the four particles, p, n, e, and γ, nature has hidden the role of all other elementary particles in the structure of normal stable matter. We know that the protons and neutrons themselves are not simple point particles but have a size and a structure. Other elementary particles are involved in this structure, but the constituents are so tightly bound (binding energies comparable to their rest mass times c^2) that it takes relatively enormous energy to break apart a proton (or a neutron). To break up a proton we must supply enough energy to create the rest mass and the kinetic energy of the resulting particles.

In an energetic neutron-proton collision we can produce neutral π^0 mesons:

$$n + p \rightarrow n + p + \pi^0,$$

provided the kinetic energy of the incident neutron is sufficiently large. It must be large enough to equal the kinetic energy of the final n, p, and π^0 plus the rest energy of the π^0 meson,

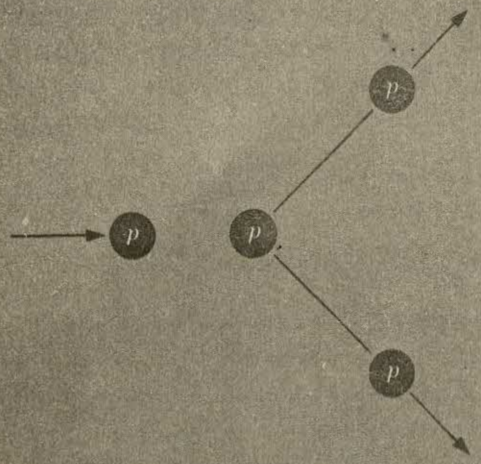
$$M(\pi^0)c^2 \cong 140 \text{ Mev}.$$

From the standpoint of energy and momentum conservation, the

π^0 was created in this collision; it did not exist at all before the collision. The neutron and proton supplied the energy and momentum to catalyze the creation of the π^0 meson. The π^0 meson can be viewed as created from the vacuum, just as electron-positron pairs are created by a γ ray. The detailed description of the mechanism of such processes is possible only in the language of relativistic quantum theory. The interaction between pions (π mesons) and nucleons (protons or neutrons) is such that if we could look within a nucleon with an ideal microscope (using very short wavelength radiation), we would detect the presence of π^0 mesons there; but if we were to knock any such meson out of a proton or a neutron, the absence would be quickly corrected in the nucleon by the creation of a new meson there. The final proton or neutron would be in every way identical to a nucleon which had not catalyzed the creation of a meson. Because the *net* result is the creation of a free meson (the one knocked out of the nucleon), energy conservation requires that all the rest mass energy of the meson must be supplied in the reaction.

In order to examine most of the elementary particles, all except the few stable ones (electrons, protons, neutrons bound in nuclei, photons), it is necessary to create them in energetic collisions. Even then the particles may decay so rapidly that we must continually create them anew whenever we wish to explore any of their properties. Before 1949, when the great accelerators began to operate, the creation of unstable particles could only be observed from the collisions of high-energy cosmic rays (generally protons or atomic nuclei) with the nuclei in the atmosphere or in some dense material set up as a target. Unfortunately, the small numbers and uncontrollable energies of the incident particles made it extraordinarily difficult to acquire accurate data concerning their properties. Nevertheless, many very significant data have been acquired from cosmic-ray experiments.

To create an elementary particle of mass M takes energy at least sufficient to exceed the rest mass energy Mc^2. This is not very much; the heaviest elementary particles now known are only 4000 times as massive as an electron, so that their rest energy is at most a few thousandths of an erg. A flashlight battery can supply enough energy to create thousands of particles per second. The problem is to localize this energy sufficiently, to put this much energy into the very small volume ($\sim 10^{-40}$ cm^3) of space that is occupied by a single particle. This is accomplished with a great accelerator that initiates a collision in which a single, incoming



Low-energy collision of two protons results merely in elastic scattering. The two protons are very much like billiard balls.

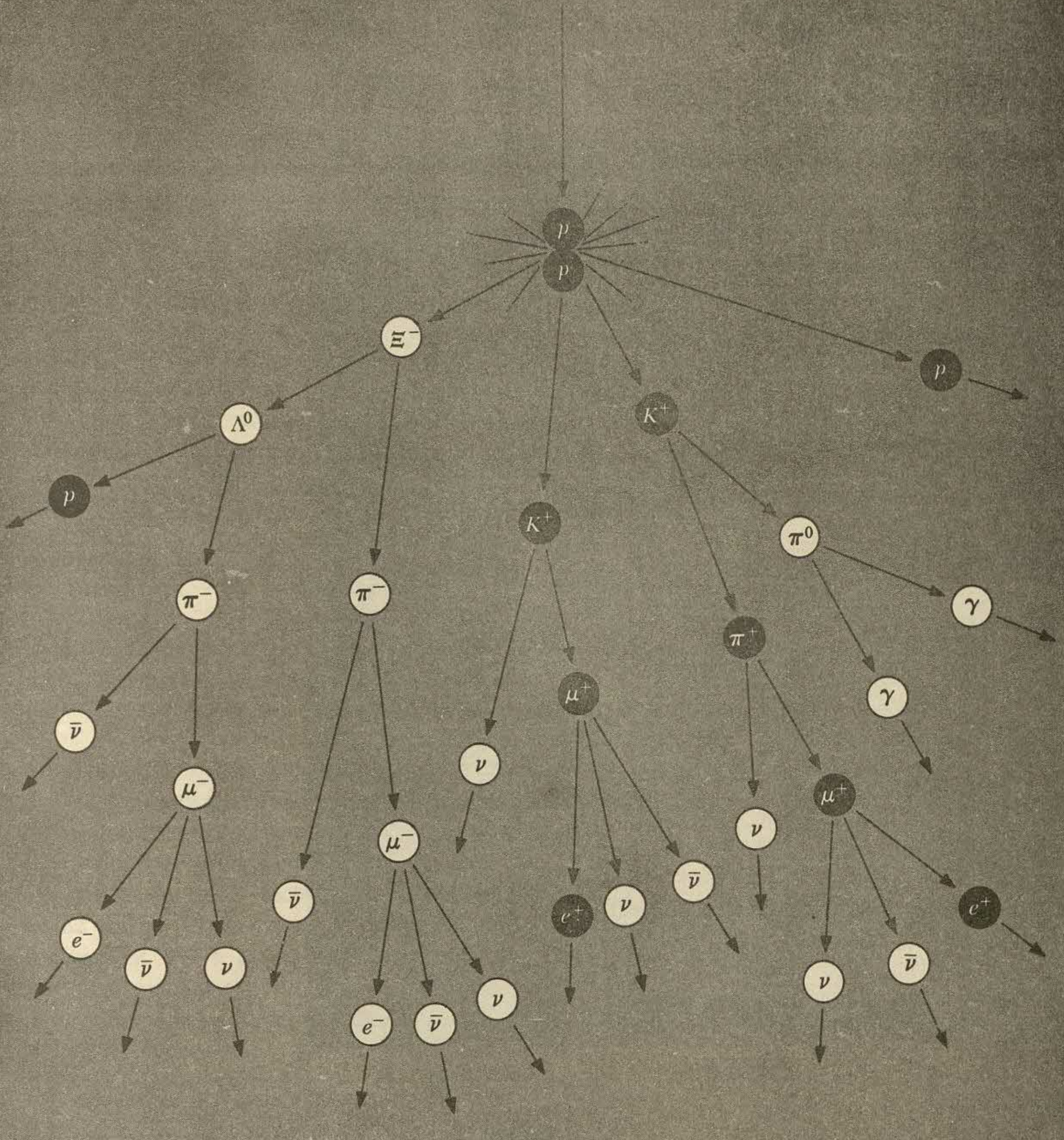
particle brings in enough energy by itself to start a desired reaction or to create one or more elementary particles. The high-energy accelerators are chiefly used to accelerate protons, but present high-energy electron accelerators are used when the object is to study the structure of the proton or neutron.

The world of elementary particle physics has largely been created by the physicists who have explored it. In the study of an unstable particle, the particle must be detected quickly and the relevant measurements recorded before it decays or is absorbed. In the table we give a compilation (made in part by W. H. Barkas and A. H. Rosenfeld) of the masses and mean lives of many of the more stable of the elementary particles.

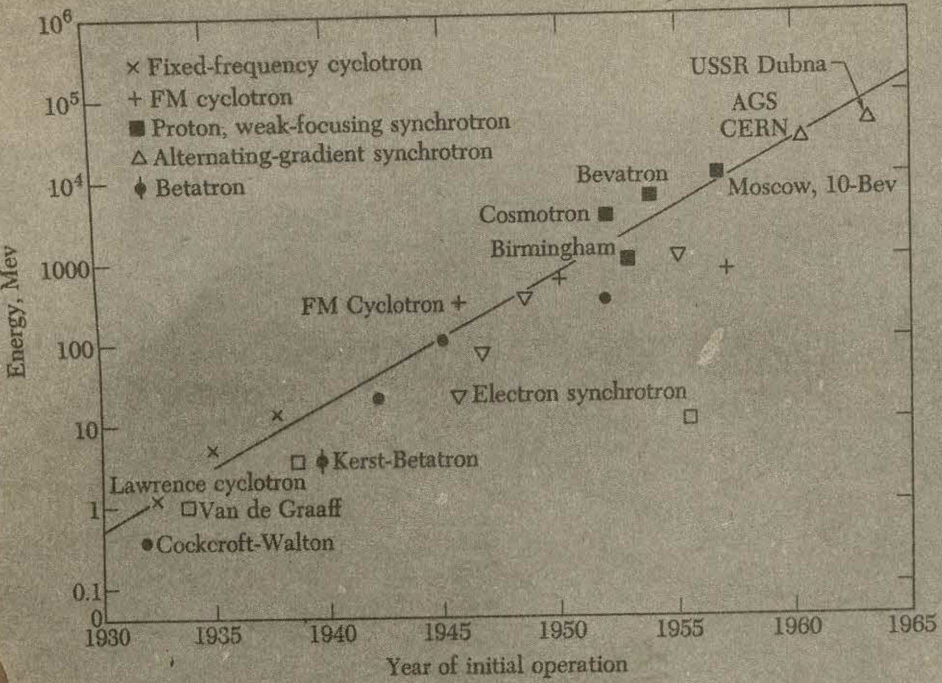
Masses and mean lives of particles

(The antiparticles, as far as we know, have the same spins, masses, and mean lives as the particles listed, but opposite charge)

Particle	*Spin*	*Mass, Mev*	*Mean life, sec*
Photon			
γ	1	0	Stable
Leptons			
ν (two kinds)	$\frac{1}{2}$	0	Stable
$e^{\mp}$	$\frac{1}{2}$	0.511006 ± 0.000005	Stable
$\mu^{\mp}$	$\frac{1}{2}$	105.655 ± 0.010	2.212×10^{-6}
Mesons			
π^+	0	139.59 ± 0.05	2.55×10^{-8}
π^0	0	135.00 ± 0.05	$(2.2 \pm 0.8) \times 10^{-16}$
$K^{\pm}$	0	493.9 ± 0.2	1.22×10^{-8}
K^0	0	497.8 ± 0.6	50% K_1, 50% K_2
K_1			1.00×10^{-10}
K_2			6×10^{-8}
Baryons			
p	$\frac{1}{2}$	938.256 ± 0.015	Stable
n	$\frac{1}{2}$	939.550 ± 0.015	1.01×10^{3}
Λ	$\frac{1}{2}$	1115.36 ± 0.14	2.5×10^{-10}
Σ^+	$\frac{1}{2}$	1189.40 ± 0.20	0.8×10^{-10}
Σ^-	$\frac{1}{2}$	1197.4 ± 0.30	1.6×10^{-10}
Σ^0	$\frac{1}{2}$	1193.0 ± 0.5	$<0.1 \times 10^{-10}$
Ξ^-	?	1318.4 ± 1.2	1.3×10^{-10}
Ξ^0	?	1311 ± 8	1.5×10^{-10} (1 event)
Ω^-		1676	



High-energy collision of two protons produces a shower of particles. One of the original protons darts away unchanged (*upper right*). The other gives rise to a negative xi particle and two positive *K* mesons. These particles, being unstable, then decay into other particles, some of which in turn decay. At the end only stable particles remain, some twenty of them in the event depicted here. (*S. B. Treiman, "The Weak Interactions," Sci. American* **200,** 77, *March 1959*)



Energies of high-energy particle accelerators throughout the world have increased approximately by a factor of 10 every five years since 1932. Each major jump has been initiated by some fundamental new idea in the design of accelerators. The first fixed-frequency cyclotron was the Lawrence cyclotron in 1932; the first frequency-modulated cyclotron was the Berkeley 184-in. machine in 1946; the first proton weak-focusing synchrotron was the cosmotron in 1952; the alternating-gradient synchrotron originated with the CERN and Brookhaven machines in 1959 and 1960. (*After R. D. Hill, "Tracking Down Particles," W. A. Benjamin, Inc., New York, 1963*)

Masses. An elementary particle always has the same rest mass.† If two particles have different rest masses, we consider them different particles. The magnitude of the rest mass serves as the principal label which identifies the particle uniquely. The value of the rest mass may allow us to infer the existence of a particle even without a direct observation, just from the conservation of energy and momentum. Charged particles leave evidence of their trajectories in appropriate material (cloud chambers, bubble chambers, photographic plates, counters). With a magnetic field we can measure the momentum and hence the energy of the particles before and after the collision.

---

† This is exactly true only for the stable particles. For unstable particles, experiment confirms the quantum theory prediction that the mass M has an intrinsic uncertainty ΔM given by $(\Delta M)c^2 \sim h/2\pi\tau$ with τ the mean life and h Planck's constant. For the neutron, $\tau \sim 15$ min and $(\Delta M/M) \sim 10^{-27}$. For the π^0, $\tau \sim 10^{-16}$ sec and $(\Delta M/M) \sim 10^{-7}$. For the very short-lived particles with $\tau \ll 10^{-16}$ sec, the mass spread is no longer small.

As an example, consider the collision of a high-energy electron with a proton at rest. The proton might be in a hydrogen molecule in a target of liquid hydrogen. In most collisions the incident electron plus a particular proton form nearly an isolated system, and very little momentum or energy is transferred to any other particle. If the unprimed variables refer to the particles before the collision and the primed variables refer to them after the collision,

$$\mathbf{p}_1 + \mathbf{p}_2 - \mathbf{p}_1' - \mathbf{p}_2' \equiv \Delta\mathbf{p} = 0, \tag{1}$$

and

$$[p_1^2c^2 + m^2c^4]^{1/2} + [p_2^2c^2 + M_p^2c^4]^{1/2} - [p_1'^2c^2 + m^2c^4]^{1/2} - [p_2'^2c^2 + M_p^2c^4]^{1/2} \equiv \Delta E = 0 \tag{2}$$

Here M_p is the proton rest mass and m is the electron rest mass. However in about one percent of the collisions, it may be observed that ΔE and $\Delta\mathbf{p}$ are not zero. This is what we would expect if an uncharged particle (which leaves no visible evidence of its trajectory) is produced in these electron-proton collisions. But if momentum and energy were not always conserved, a similar situation would also exist. How do we distinguish between the two possibilities? A significant experimental result is that in all the collisions with different magnitudes and directions for $\mathbf{p}_1'$, $\mathbf{p}_2'$, the missing energy is always positive. If it were ever found to be negative, we should not be able to maintain that the missing energy was converted into rest mass and kinetic energy of unobserved particles. More important, as long as only a single invisible (neutral) particle of mass M is created which carries away ΔE and $\Delta\mathbf{p}$, the two quantities must always be related by

$$(\Delta E)^2 - c^2(\Delta\mathbf{p})^2 = M^2c^4 \geqq 0. \tag{3}$$

The rest mass M of the neutral particle must be independent of $\mathbf{p}_1$, $\mathbf{p}_1'$, $\mathbf{p}_2$, and $\Delta\mathbf{p}$. In the millions of inelastic collisions that have been analyzed, the missing energy and momentum have always been successfully interpreted in terms of the creation of one or more of the known elementary particles. This provides perhaps the strongest existing evidence for the conservation of relativistic momentum and energy.

In electron-proton collisions, so long as the kinetic energy of the incident proton is less than 140 Mev, almost all inelastic collisions ($\Delta E \neq 0$) give

$$(\Delta E)^2 - c^2(\Delta p)^2 = 0. \tag{4}$$

This, by comparison with (3), means that the particle created in the collision has zero rest mass. The particle is identifiable as a photon γ, so that the reaction is

$$e + p \rightarrow e + p + \gamma. \tag{5}$$

In the electron energy range below 140 Mev we can consistently interpret those instances in which $(\Delta E)^2 > c^2(\Delta p)^2$ in terms of multiple γ production processes, such as $e + p \rightarrow e + p + \gamma + \gamma$. No collisions have ever been seen in which $(\Delta E)^2 < c^2(\Delta p)^2$.

At incident electron energies above approximately 140 Mev another mechanism of energy-momentum loss occurs. In this range it is observed in inelastic collisions that

$$(\Delta E)^2 - c^2(\Delta p) = (135 \text{ Mev})^2. \tag{6}$$

We interpret this energy-momentum imbalance as the creation of a particle of rest mass 135 Mev. The missing energy and momentum in (6) are given to a neutral π^0 meson of rest mass $264m$, where m is the mass of an electron. The process is

$$e + p \rightarrow e + p + \pi^0.$$

This reaction has been confirmed by observation of the decay products of the unstable π^0, which decays according to

$$\pi^0 \rightarrow \gamma + \gamma.$$

This is how the neutral pion was first discovered. However, repeated verification of (6) for varied incident energies, ΔE, and Δp is in itself sufficient to confirm the existence of a meson of rest mass $264m$. The π^0 is also produced in high-energy collisions of other particles; for example, in the processes

$$\gamma + p \rightarrow p + \pi^0, \qquad n + p \rightarrow n + p + \pi^0. \tag{7}$$

In an analogous way the discovery and confirmation of other elementary particles is often based upon the application of the laws of conservation of energy and momentum.

Charge. All known elementary particles have charge $+e$, $-e$, or zero, within the accuracy of the measurements. Further, no example has ever been reported of a collision process in which the charge is not conserved, within the experimental accuracy. For example, from the absence of deflection of a beam of neutrons in a uniform electric field, the charge of the neutron is known to be zero within 10^{-17} of the electron charge.

If we apply charge conservation as a principle to the reactions

$$n + p \rightarrow p + p + \pi^-; \qquad n + p \rightarrow n + n + \pi^+, \tag{8}$$

the charge of the π^- and π^+ mesons is determined as $-e$ and $+e$, respectively. Other particles are produced in positive and negative pairs, as in

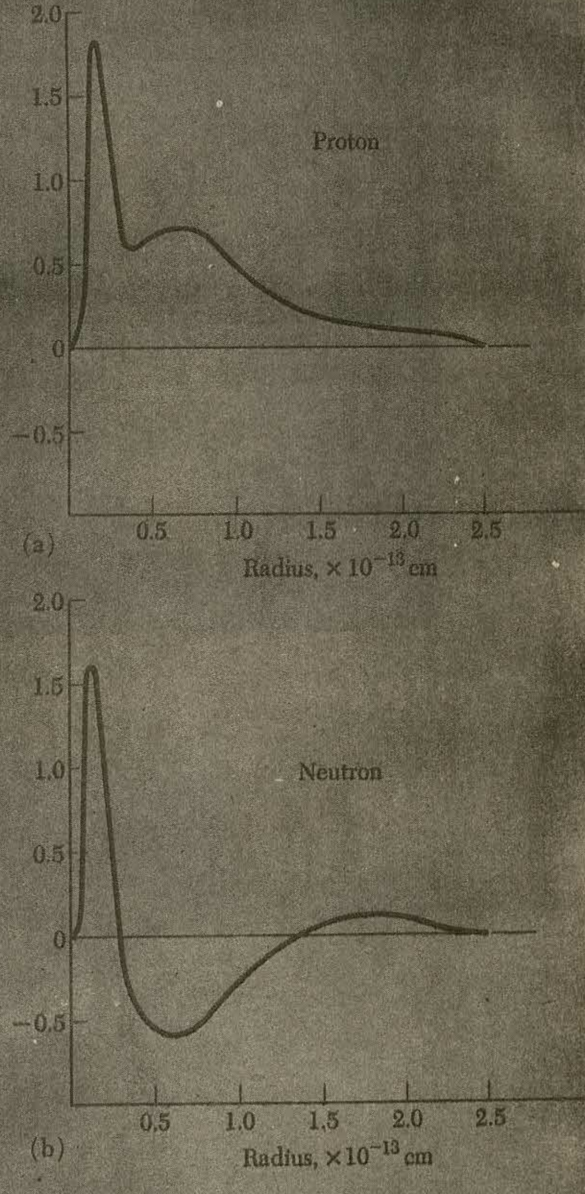
$$n + p \rightarrow n + p + K^- + K^+. \tag{9}$$

Here charge conservation says that the charges of K^- and K^+ should be equal in magnitude but opposite in sign.

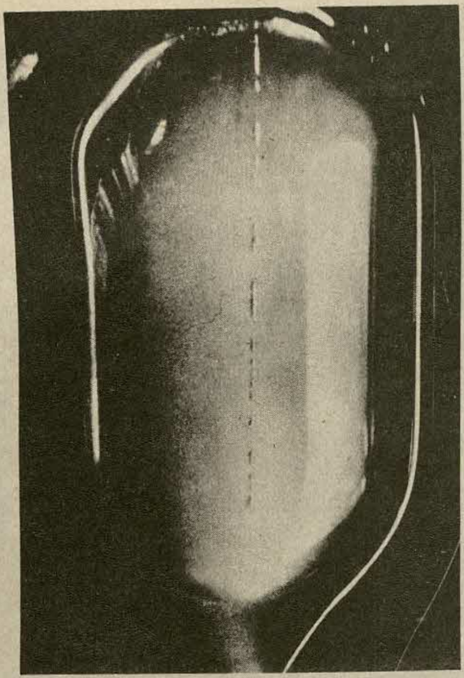
We believe that the charges of all elementary particles are very accurately $\pm e$, or 0. Independent evidence for charge conservation and for the quantization of the charges of elementary particles comes indirectly from the fact that the value of the charge q enters into the determination of the momentum p of a particle, based upon measurement of the radius ρ of the circle the particle makes in a magnetic field $\mathbf{B}$:

$$B\rho = \frac{cp}{q}. \tag{10}$$

This relation was derived relativistically in Chap. 13. Such measurements are the most important means of determining momenta and of verifying the energy-momentum relation, Eq. (3). Departure of the charge of a particle from its assumed value $\pm e$ would change both the computed momentum and energy of the particle. This in turn would give rise to an apparent lack of energy-momentum conservation in reactions. Generally, however, Eq. (3) cannot be verified with great precision in collisions which produce unstable particles, because momenta can rarely be determined by curvature measurements to better than 1 percent. Only for the stable elementary particles—the proton, neutron, electron, photon, and neutrino—does there exist very precise direct confirmation of the assumption that the charges are $\pm e$ or 0, to better than $10^{-15}e$.

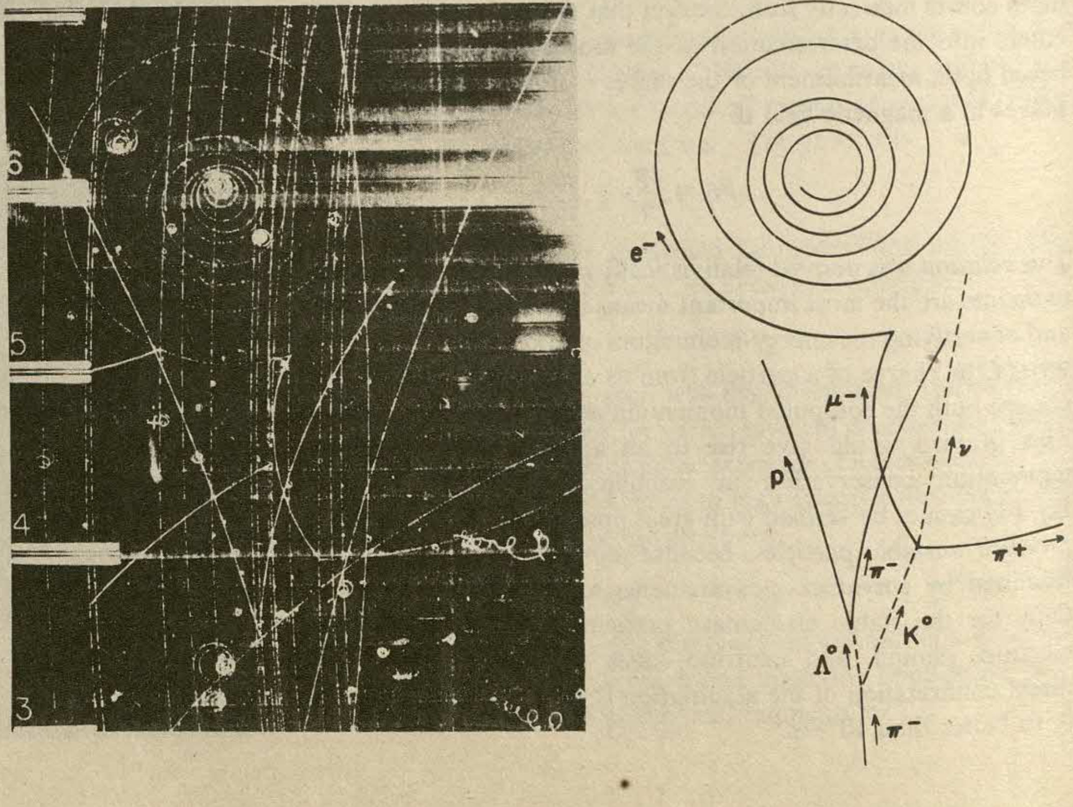


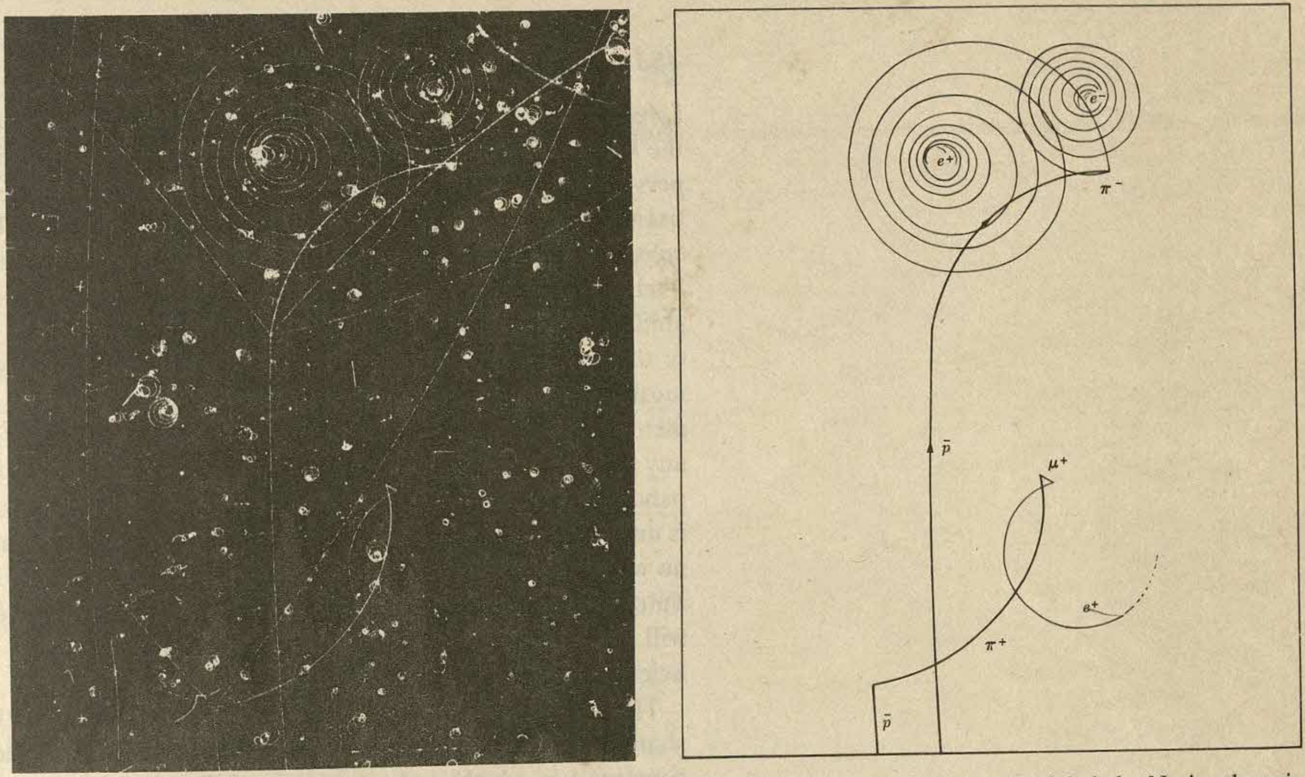
Distributions of charge in the proton and neutron as functions of radius within the particle. The ordinate is proportional to the charge in a thin spherical shell of radius r. The area under the proton curve is equal to the proton charge. The area under the neutron curve is equal to zero. The data are obtained by high-energy electron-scattering experiments.



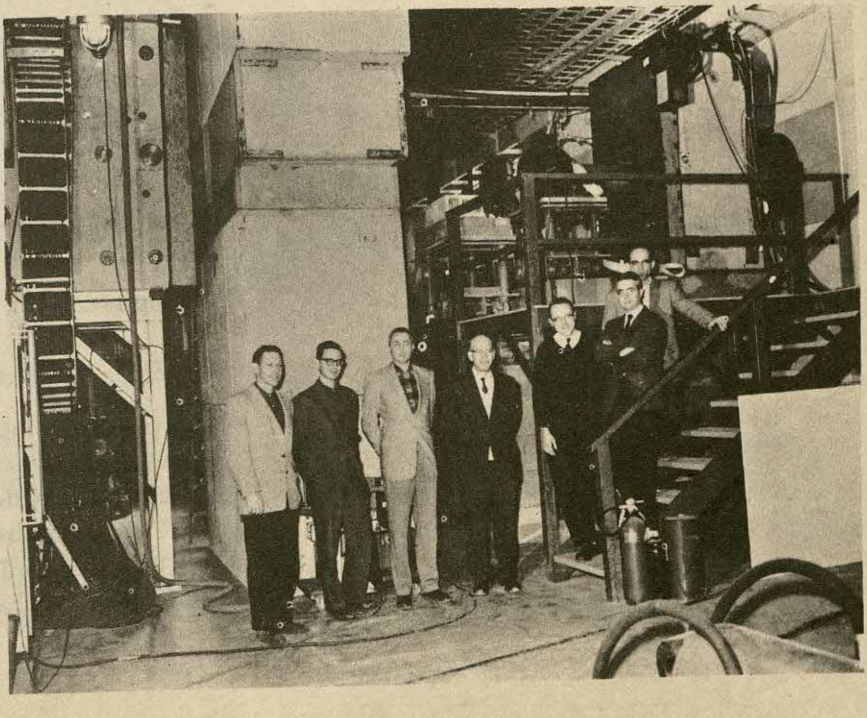
An early bubble chamber, showing tracks of cosmic-ray particles in diethyl ether. This was constructed at the University of Michigan. (*Courtesy of D. Glaser*)

Mesons and friends in a hydrogen bubble chamber. (*Lawrence Radiation Laboratory*)





Particle tracks in a liquid hydrogen bubble chamber. We have labeled two antiprotons which enter from the bottom. Both of these recombine with protons in the chamber. Several reaction products are also labeled. Notice the spirals made by electrons and positrons. (*Lawrence Radiation Laboratory*)



A representative group of the individuals involved in the observation of the omega minus photographed at Brookhaven's 80-in. bubble chamber: *from left to right,* Jack E. Jensen, Nicholas P. Samios, William A. Tuttle, Ralph P. Shutt, Medford S. Webster, William B. Fowler, and Donald P. Brown. (*Brookhaven National Laboratory*)

Lifetime. The decay of an unstable particle is very different from the kind of disintegration we commonly see around us. An aged person is more likely to die in the next hour than a young person; a bacterium will not undergo fission immediately after birth, but will split only after a specified time has elapsed; an old car is more likely to break down than a new one. In these cases the probability of some type of disruption depends also upon the history of the object up to that time: those which have lived longest are most subject to some process of disintegration. But it is an experimental fact that the decay probability of elementary particles or any kind of radioactive nuclei or excited atom or molecule is independent of the length of time the particle has lived. A free neutron is unstable, but a neutron which has already lived a long time is in no way different from a neutron which has been free only a short time. It is impossible to predict when a given unstable particle will decay. Only an *average* lifetime for a large number of particles has a reproducible meaning.

The probability $P_{\Delta t}$ that a particle at time t will decay in the next short interval of time Δt is equal to the product of Δt and some constant $1/\tau$, which is characteristic of the particle but not of its history:

$$P_{\Delta t} = \frac{\Delta t}{\tau}. \qquad (11)$$

By a short interval we mean that $\Delta t \ll \tau$. For a large number of particles N, the number which decay in Δt is $NP_{\Delta t}$. This decay changes the number of particles by $-\Delta t(dN/dt)$. Therefore, using (11),

$$NP_{\Delta t} = \frac{N\,\Delta t}{\tau} = -\Delta t\frac{dN}{dt}; \qquad \frac{dN}{dt} = -\frac{N}{\tau}. \qquad (12)$$

The solution of (12) is

$$N(t) = N_0 e^{-t/\tau}, \qquad (13)$$

where N_0 is the number of particles at $t = 0$.

The quantity τ is called the *mean life* of the unstable particle; it is conventionally defined in the rest frame of the decaying particle. The probabilistic nature of decay processes as expressed by (13) is in accord both with experiment and with quantum mechanical theoretical explanations.

Many unstable particles have more than one decay mode. Consider the two modes of decay of a Λ particle (neutral):

$$\text{Mode (A):} \qquad \Lambda \rightarrow p + \pi^-; \tag{14}$$

$$\text{Mode (B):} \qquad \Lambda \rightarrow n + \pi^0. \tag{15}$$

The probability for any decay in the next Δt is the sum of the probability for decay through modes (A) and (B):

$$P = P_A + P_B; \qquad P_A = \frac{1}{\tau_A}\Delta t; \qquad P_B = \frac{1}{\tau_B}\Delta t. \tag{16}$$

Here

$$-\Delta t \frac{dN}{dt} = PN \tag{17}$$

or

$$\frac{dN}{dt} = -\frac{1}{\tau}N, \tag{18}$$

with

$$\frac{1}{\tau} = \frac{1}{\tau_{\mathrm{A}}} + \frac{1}{\tau_{\mathrm{B}}}. \tag{19}$$

We see that the number of undecayed Λ remaining at any time t is again given by a simple exponential, as in (13). If we were to detect only π^- mesons emitted from a sample of Λ particles, we would have $P_{\mathrm{A}}N$ for the number of π^- mesons emitted in time Δt. Then the rate R^- of π^- emission is given by

$$R^- = \frac{N_0}{\tau_{\mathrm{A}}} e^{-t/\tau}. \tag{20}$$

The rate R^0 of π^0 emission is given by

$$R^0 = \frac{N_0}{\tau_{\mathrm{B}}} e^{-t/\tau}. \tag{21}$$

The total decay rate R is

$$R = R^- + R^0 = \frac{N_0}{\tau} e^{-t/\tau}, \tag{22}$$

where we have used (19). No matter which decay mode is observed, the exponential function is the same. A particle has only one single mean life, no matter how many different ways it has to decay and no matter which decay mode is used to measure the exponential decay.

Except for the electrostatic interaction between charged particles, the forces between elementary particles become small when their separation exceeds 2×10^{-13} cm. Therefore, even when the decay products of an unstable particle travel at the velocity of light, it generally takes about $(2 \times 10^{-13}/c) \approx 10^{-23}$ sec before a decay can be considered irrevocable and complete.

A particle which decays in a time comparable to 10^{-23} sec would hardly be called a particle. It would take this long for the decay products to separate even if they were never bound into a particle at all. This time interval, 10^{-23} sec, gives a time standard relative to which we can in some sense consider decays to be fast or slow. From the table we see that all the decays listed (except π^0 and Σ^0 which can involve just the emission of a photon) are exceedingly slow in comparison with 10^{-23} sec, with mean lives varying between 17 min for the neutron and 10^{-10} sec for Λ or $\Sigma^{\pm}$. Generally the more kinetic energy available for the decay products, the more rapid is the decay. Even a "long-lived" particle of $\sim 10^{-10}$ sec lives such a short time compared to those times which allow convenient laboratory measurements that the problem of studying the properties of these unstable elementary particles demands special methods, apparatus, and ingenuity.

Other Characteristics. A knowledge of mass, charge, and lifetime does not exhaust the measurable properties of elementary particles any more than these properties give a complete description of ordinary macroscopic objects. They are usually the earliest to be measured, however, and generally serve as a unique signature for a particular elementary particle.

Most elementary particles can be classified into special groupings (multiplets). For example, there are three pions: π^+, π^0, π^- of nearly the same mass (see table) and with strikingly similar properties except for charge. But there is no π^{++} or π^{--}, etc.; the pions occur in triplets. Nucleons on the other hand occur in doublets, p and n (or rather a pair of doublets, because of the existence of the antiproton $\overline{p}$ and the antineutron $\overline{n}$). Again, there is no multicharged nucleon. Similarly the Ξ particle also occurs as a pair of doublets. So do the K particles. But Λ is a singlet, and Σ^+, Σ^0, Σ^- form a triplet.

Particles of the mass of the muons and lighter (μ, e, ν) are called *leptons* (light particles); particles intermediate in mass between leptons and nucleons (π, K) are called *mesons* (in-between particles), and the nucleons and heavier particles (p, n, Λ, Σ^+, Σ^0, Σ^-, Ξ^0, Ξ^-, Ω^-) are called *baryons* (heavy particles). This broad classification is very convenient in discussing the interactions among elementary particles. (For example, in *all* reactions the number of baryons *minus* the number of antibaryons is never changed. This is not true for mesons but is apparently valid for leptons.)

Many elementary particles spin in a manner analogous to that of the earth on its axis, but with certain differences which originate in the fact that this spinning motion of elementary particles must be described by quantum mechanics. Here we shall merely summarize a result: all leptons and long-lived baryons seem to have the same intrinsic angular momentum $\frac{1}{2}\hbar$, while long-lived mesons have none at all.

On the average the charge of an elementary particle is distributed throughout the particle. In any "gentle" experiment which does not in itself disrupt the particle, only this time average is measurable since the measurement cannot be instantaneous. (Again, it is quantum mechanics which limits our ability to describe the structure of an elementary particle.) The experimental determination of this charge distribution for the proton, neutron, and electron gives important confirmation to the model of the electron as a point charge (at least down to 10^{-14} cm), while the proton and neutron appear as more-complicated structures with charge distributed within a sphere of radius of approximately 10^{-13} cm. For leptons the magnetic moments (to be defined in Vol. II) become larger as the inverse of the mass, except that ν and γ have no detectable intrinsic magnetic moments. Not only the magnitude of the magnetic field but also the exact distribution of the currents which give rise to it can, in principle, be measured. One of the greatest accomplishments of relativistic quantum theory is the successful prediction of the observed intrinsic magnetic field of the electron to within 1 part in 10^5, less than the present experimental uncertainty.

Four Basic Forces of Nature. We know of only four basic ways in which matter can interact. That is, there are four fundamental interactions which account for all the known forces in the universe:

gravitational interaction
electromagnetic interaction
strong interaction
weak interaction

The gravitational interaction, which is the weakest of all, is the interaction that holds the earth together, binds the sun and planets into the solar system, and binds stars into galaxies. It is responsible for the large-scale drama of the universe.

The electromagnetic interaction binds electrons to atoms and binds atoms together to form molecules and crystals. It is the significant interaction for all of chemistry and biology.

The strong interaction glues nucleons together; it binds neutrons and protons to form the nuclei of all the elements. The strongest force known in nature, it is also of quite short range. It is the dominant interaction in high-energy nuclear physics.

The weak interaction causes the light elementary particles (the leptons: electrons, neutrinos, and muons) to interact with each other and with heavier particles. The weak interaction, which is associated with the β decay of radioactive nuclei, is a very short-range interaction. The weak interaction cannot form stable states of matter in the sense that the gravitational force can form a solar system.

In Vol. IV elementary particles will be discussed in greater depth.

Further reading

K. W. Ford, *The world of elementary particles* (Blaisdell, Boston, 1963).

P. Morrison, "Neutrino astronomy," *Sci. American* **207**, 90 (August, 1962).

C. N. Yang, *Elementary particles* (Princeton Univ. Press, Princeton, N. J., 1961).

Problems

1. An alpha particle ($M_\alpha \approx 4M_p$; charge $2e$) of kinetic energy 6 Mev is emitted by a radioactive source. Suppose that the particle is directed exactly toward the center of a gold nucleus, of charge $79e$. Assume that the nuclear charge is concentrated at a point, and neglect the recoil of the nucleus.

(*a*) How close does the alpha particle come to the center?

(*b*) What happens to the alpha particle after it reaches this point of closest approach?

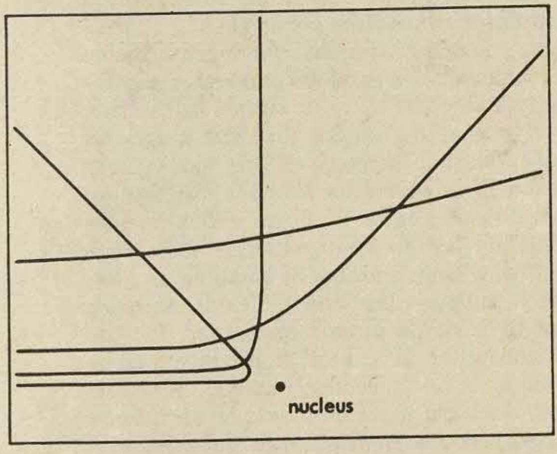
2. In the pre-Rutherford days is was thought that the charge of a nucleus was spread out over the linear dimensions of an atom, say 10^{-8} cm. Suppose that we simply omit the effect of the atomic electrons and let the alpha particle interact with a massive distribution of positive charge $79e$ of uniform density in a sphere of radius 10^{-8} cm. What is the maximum energy an alpha particle can have and still be scattered backward by such a gold nucleus? (Use the methods of Chap. 9; you will have to find an expression for the potential energy at the center of a uniformly charged sphere.)

Ans. 3400 ev.

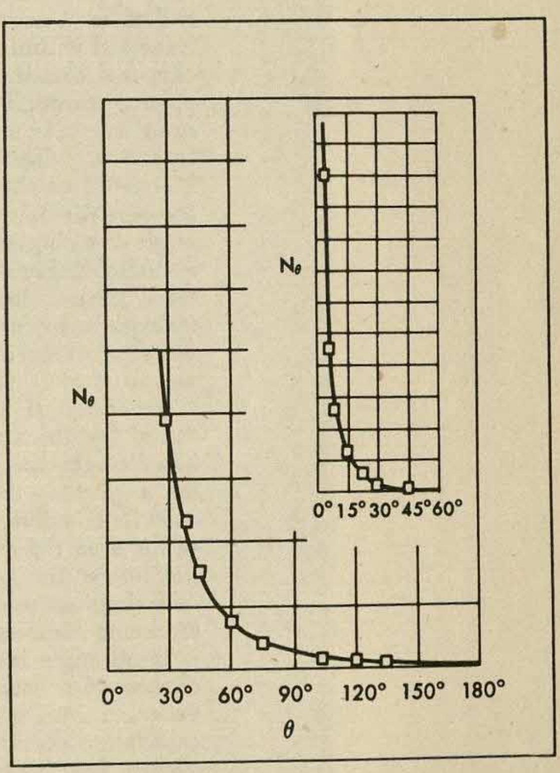
Historical Note 1. Rutherford scattering and the nuclear atom

The discovery of the nucleus was reported by Rutherford in the classic paper which we reproduce in part below [E. Rutherford, *Phil. Mag.* **21**, 669 (1911)]. The derivation of the scattering probability at large angles is quite straightforward, but utilizes trigonometric relations liberally. The result that the eccentricity of the hyperbola is sec θ (as defined in Fig. 1 below) follows directly from our Eq. (9.71) with $1 - e \cos \varphi$ in our notation set equal to zero.

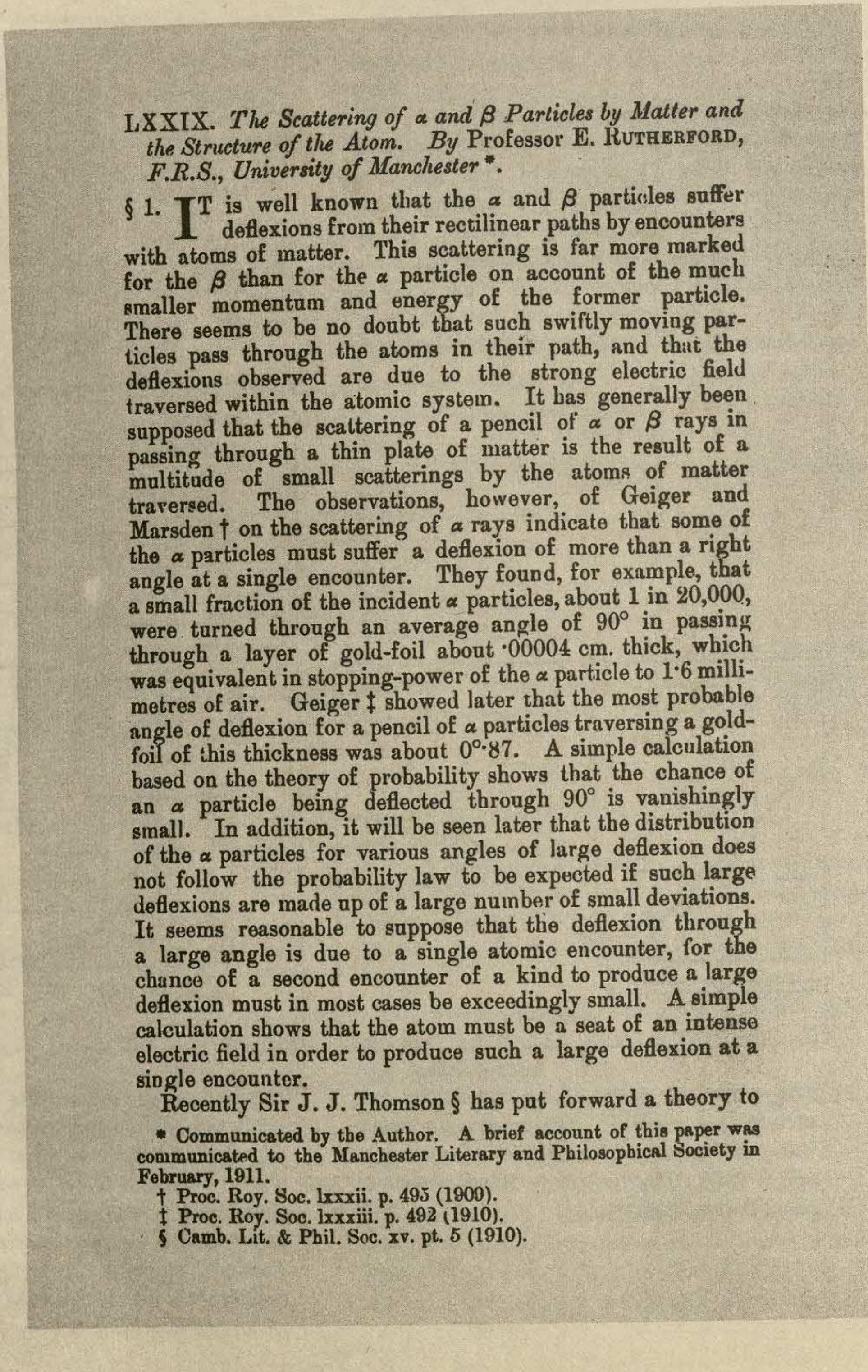
Recall that an α particle is a nucleus of He_4 and a β particle is an electron.



The computed paths of alpha particles approaching a nucleus. These were computed using an inverse-square force of repulsion. (*From PSSC, "Physics," D. C. Heath and Company, Boston, 1960*)



A graph of N_θ (the number of particles scattered through an angle greater than θ) vs θ. The curved line shows what we would expect with a Coulomb force. The points in small squares represent the data collected by Geiger and Marsden in their scattering experiments. The insert shows the curve for small angles on a different scale. (*From PSSC, "Physics," D. C. Heath and Company, Boston, 1960*)



LXXIX. *The Scattering of α and β Particles by Matter and the Structure of the Atom. By* Professor E. RUTHERFORD, *F.R.S., University of Manchester* [*].

§ 1. IT is well known that the α and β particles suffer deflexions from their rectilinear paths by encounters with atoms of matter. This scattering is far more marked for the β than for the α particle on account of the much smaller momentum and energy of the former particle. There seems to be no doubt that such swiftly moving particles pass through the atoms in their path, and that the deflexions observed are due to the strong electric field traversed within the atomic system. It has generally been supposed that the scattering of a pencil of α or β rays in passing through a thin plate of matter is the result of a multitude of small scatterings by the atoms of matter traversed. The observations, however, of Geiger and Marsden [†] on the scattering of α rays indicate that some of the α particles must suffer a deflexion of more than a right angle at a single encounter. They found, for example, that a small fraction of the incident α particles, about 1 in 20,000, were turned through an average angle of 90° in passing through a layer of gold-foil about ·00004 cm. thick, which was equivalent in stopping-power of the α particle to 1·6 millimetres of air. Geiger [‡] showed later that the most probable angle of deflexion for a pencil of α particles traversing a gold-foil of this thickness was about 0°·87. A simple calculation based on the theory of probability shows that the chance of an α particle being deflected through 90° is vanishingly small. In addition, it will be seen later that the distribution of the α particles for various angles of large deflexion does not follow the probability law to be expected if such large deflexions are made up of a large number of small deviations. It seems reasonable to suppose that the deflexion through a large angle is due to a single atomic encounter, for the chance of a second encounter of a kind to produce a large deflexion must in most cases be exceedingly small. A simple calculation shows that the atom must be a seat of an intense electric field in order to produce such a large deflexion at a single encounter.

Recently Sir J. J. Thomson [§] has put forward a theory to explain the scattering of electrified particles in passing through small thicknesses of matter. The atom is supposed to consist of a number N of negatively charged corpuscles, accompanied by an equal quantity of positive electricity uniformly distributed throughout a sphere. The deflexion of a negatively electrified particle in passing through the atom is ascribed to two causes—(1) the repulsion of the corpuscles distributed through the atom, and (2) the attraction of the positive electricity in the atom. The deflexion of the particle in passing through the atom is supposed to be small, while the average deflexion after a large number m of encounters was taken as $\sqrt{m} . \theta$, where θ is the average deflexion due to a single atom. It was shown that the number N of the electrons within the atom could be deduced from observations of the scattering of electrified particles. The accuracy of this theory of compound scattering was examined experimentally by Crowther[*] in a later paper. His results apparently confirmed the main conclusions of the theory, and he deduced, on the assumption that the positive electricity was continuous, that the number of electrons in an atom was about three times its atomic weight.

---

[*] Communicated by the Author. A brief account of this paper was communicated to the Manchester Literary and Philosophical Society in February, 1911.

[†] Proc. Roy. Soc. lxxxii. p. 495 (1909).

[‡] Proc. Roy. Soc. lxxxiii. p. 492 (1910).

[§] Camb. Lit. & Phil. Soc. xv. pt. 5 (1910).

The theory of Sir J. J. Thomson is based on the assumption that the scattering due to a single atomic encounter is small, and the particular structure assumed for the atom does not admit of a very large deflexion of an α particle in traversing a single atom, unless it be supposed that the diameter of the sphere of positive electricity is minute compared with the diameter of the sphere of influence of the atom.

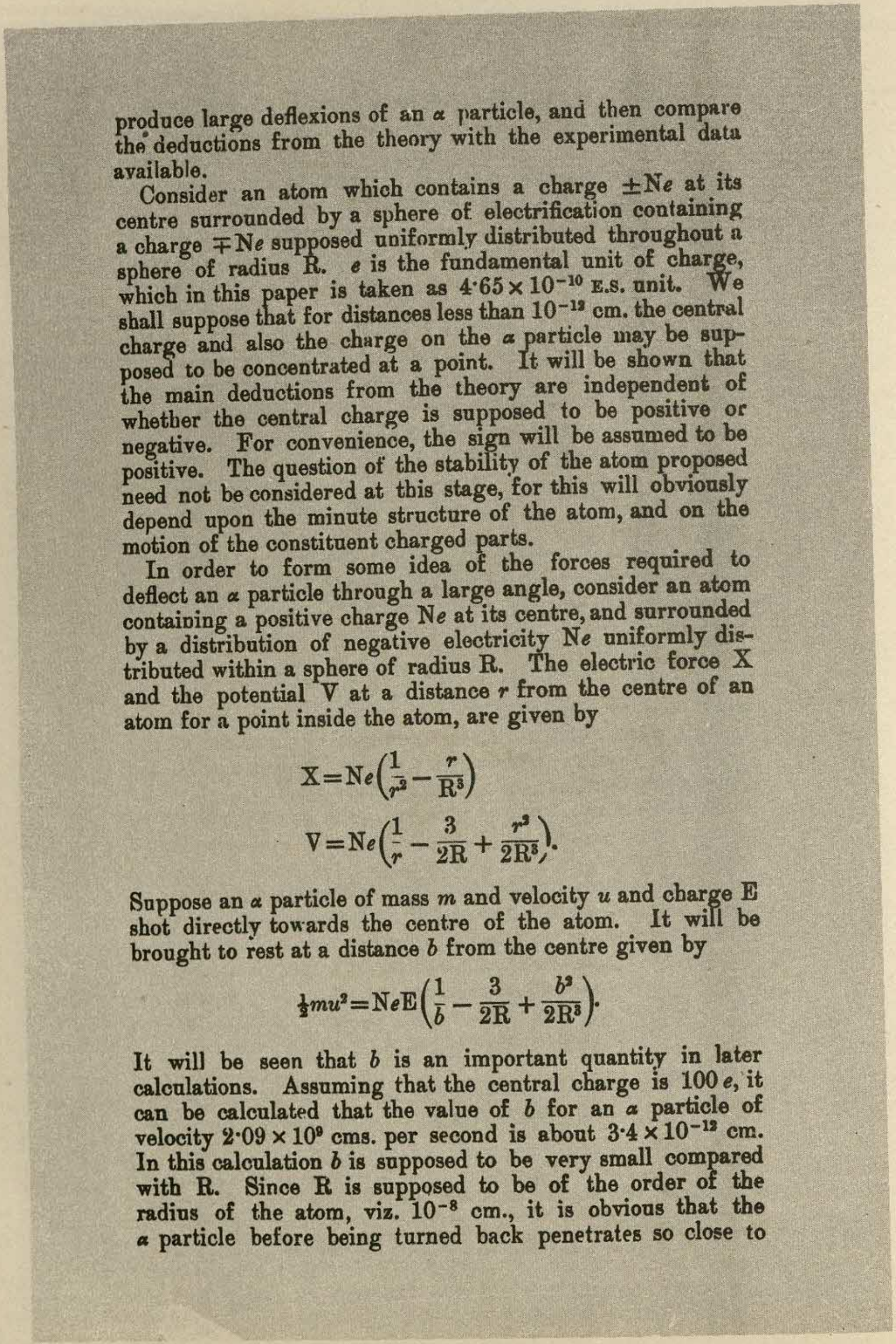
Since the α and β particles traverse the atom, it should be possible from a close study of the nature of the deflexion to form some idea of the constitution of the atom to produce the effects observed. In fact, the scattering of high-speed charged particles by the atoms of matter is one of the most promising methods of attack of this problem. The development of the scintillation method of counting single α particles affords unusual advantages of investigation, and the researches of H. Geiger by this method have already added much to our knowledge of the scattering of α rays by matter.

§ 2. We shall first examine theoretically the single encounters[†] with an atom of simple structure, which is able to

---

[*] Crowther, Proc. Roy. Soc. lxxxiv. p. 226 (1910).

[†] The deviation of a particle throughout a considerable angle from an encounter with a single atom will in this paper be called "single" scattering. The deviation of a particle resulting from a multitude of small deviations will be termed "compound" scattering.



produce large deflexions of an α particle, and then compare the deductions from the theory with the experimental data available.

Consider an atom which contains a charge $\pm Ne$ at its centre surrounded by a sphere of electrification containing a charge $\mp Ne$ supposed uniformly distributed throughout a sphere of radius R. e is the fundamental unit of charge, which in this paper is taken as $4{\cdot}65 \times 10^{-10}$ E.S. unit. We shall suppose that for distances less than 10^{-12} cm. the central charge and also the charge on the α particle may be supposed to be concentrated at a point. It will be shown that the main deductions from the theory are independent of whether the central charge is supposed to be positive or negative. For convenience, the sign will be assumed to be positive. The question of the stability of the atom proposed need not be considered at this stage, for this will obviously depend upon the minute structure of the atom, and on the motion of the constituent charged parts.

In order to form some idea of the forces required to deflect an α particle through a large angle, consider an atom containing a positive charge Ne at its centre, and surrounded by a distribution of negative electricity Ne uniformly distributed within a sphere of radius R. The electric force X and the potential V at a distance r from the centre of an atom for a point inside the atom, are given by

$$X = Ne\left(\frac{1}{r^2} - \frac{r}{R^3}\right)$$

$$V = Ne\left(\frac{1}{r} - \frac{3}{2R} + \frac{r^2}{2R^3}\right).$$

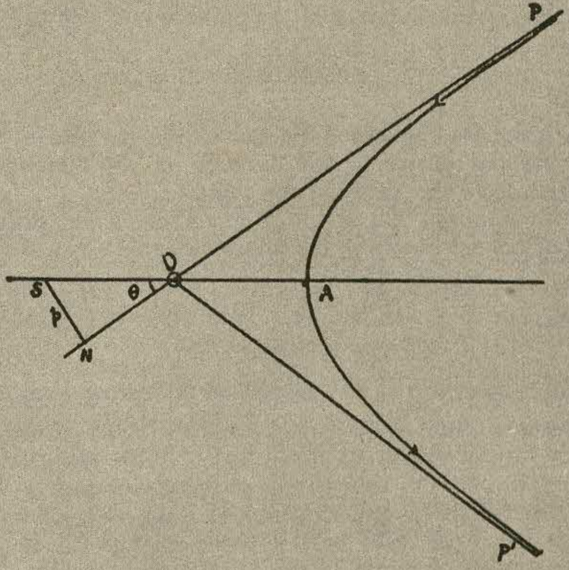
Suppose an α particle of mass m and velocity u and charge E shot directly towards the centre of the atom. It will be brought to rest at a distance b from the centre given by

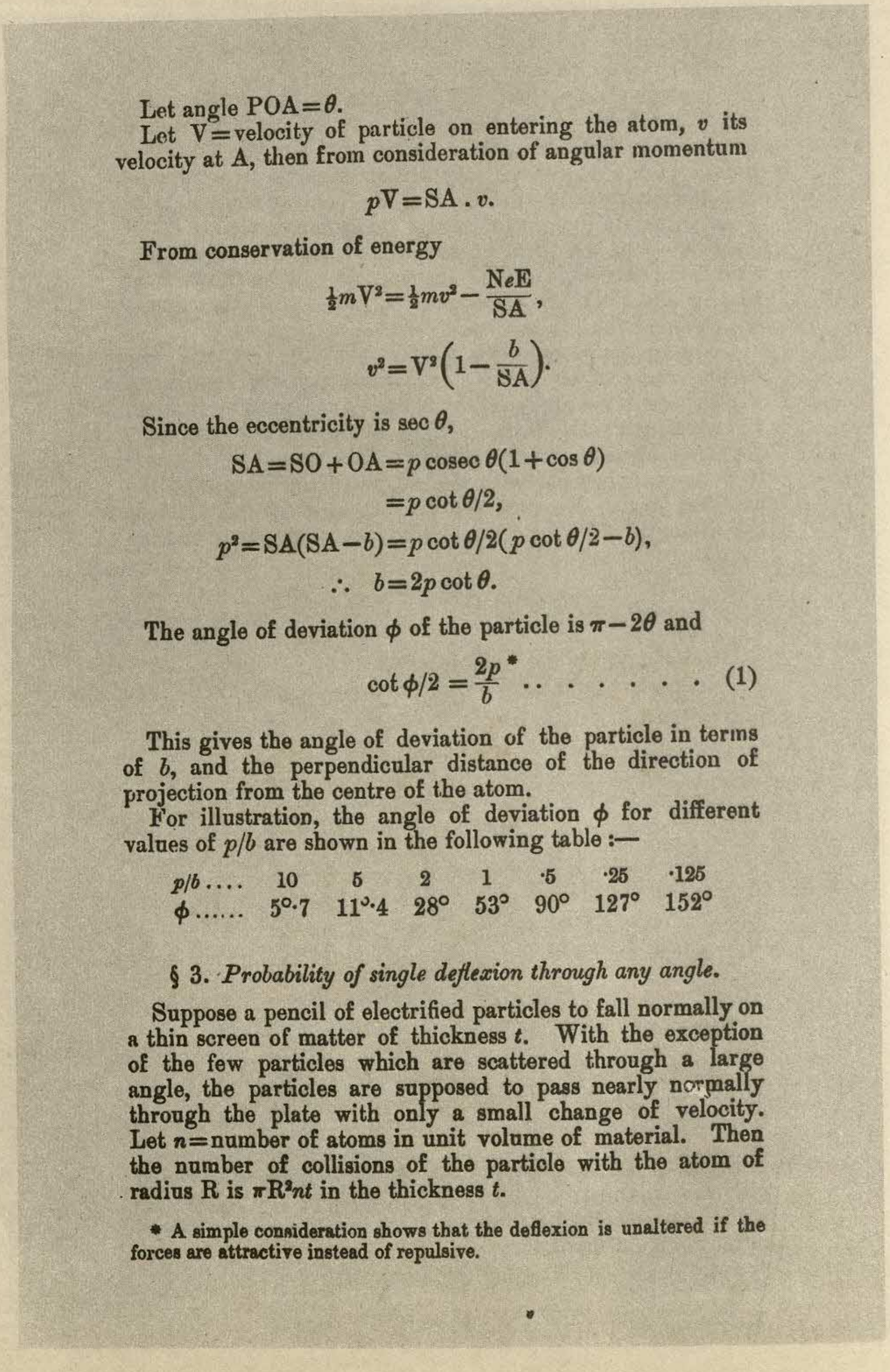
$$\tfrac{1}{2}mu^2 = NeE\left(\frac{1}{b} - \frac{3}{2R} + \frac{b^2}{2R^3}\right).$$

It will be seen that b is an important quantity in later calculations. Assuming that the central charge is $100\,e$, it can be calculated that the value of b for an α particle of velocity $2{\cdot}09 \times 10^9$ cms. per second is about $3{\cdot}4 \times 10^{-12}$ cm. In this calculation b is supposed to be very small compared with R. Since R is supposed to be of the order of the radius of the atom, viz. 10^{-8} cm., it is obvious that the α particle before being turned back penetrates so close to the central charge, that the field due to the uniform distribution of negative electricity may be neglected. In general, a simple calculation shows that for all deflexions greater than a degree, we may without sensible error suppose the deflexion due to the field of the central charge alone. Possible single deviations due to the negative electricity, if distributed in the form of corpuscles, are not taken into account at this stage of the theory. It will be shown later that its effect is in general small compared with that due to the central field.

Consider the passage of a positive electrified particle close to the centre of an atom. Supposing that the velocity of the particle is not appreciably changed by its passage through the atom, the path of the particle under the influence of a repulsive force varying inversely as the square of the distance will be an hyperbola with the centre of the atom S as the external focus. Suppose the particle to enter the atom in the direction PO (fig. 1), and that the direction of motion on escaping the atom is OP′. OP and OP′ make equal angles with the line SA, where A is the apse of the hyperbola. p=SN=perpendicular distance from centre on direction of initial motion of particle.

Fig. 1.





Let angle POA$=\theta$.

Let V=velocity of particle on entering the atom, v its velocity at A, then from consideration of angular momentum

$$pV = SA\,.\,v.$$

From conservation of energy

$$\tfrac{1}{2}mV^2 = \tfrac{1}{2}mv^2 - \frac{NeE}{SA},$$

$$v^2 = V^2\left(1 - \frac{b}{SA}\right).$$

Since the eccentricity is $\sec\theta$,

$$SA = SO + OA = p\,\mathrm{cosec}\,\theta(1+\cos\theta)$$
$$= p\cot\theta/2,$$
$$p^2 = SA(SA-b) = p\cot\theta/2(p\cot\theta/2 - b),$$
$$\therefore\ b = 2p\cot\theta.$$

The angle of deviation ϕ of the particle is $\pi - 2\theta$ and

$$\cot\phi/2 = \frac{2p}{b}\ ^{*} \ \dots\dots\dots \quad (1)$$

This gives the angle of deviation of the particle in terms of b, and the perpendicular distance of the direction of projection from the centre of the atom.

For illustration, the angle of deviation ϕ for different values of p/b are shown in the following table:—

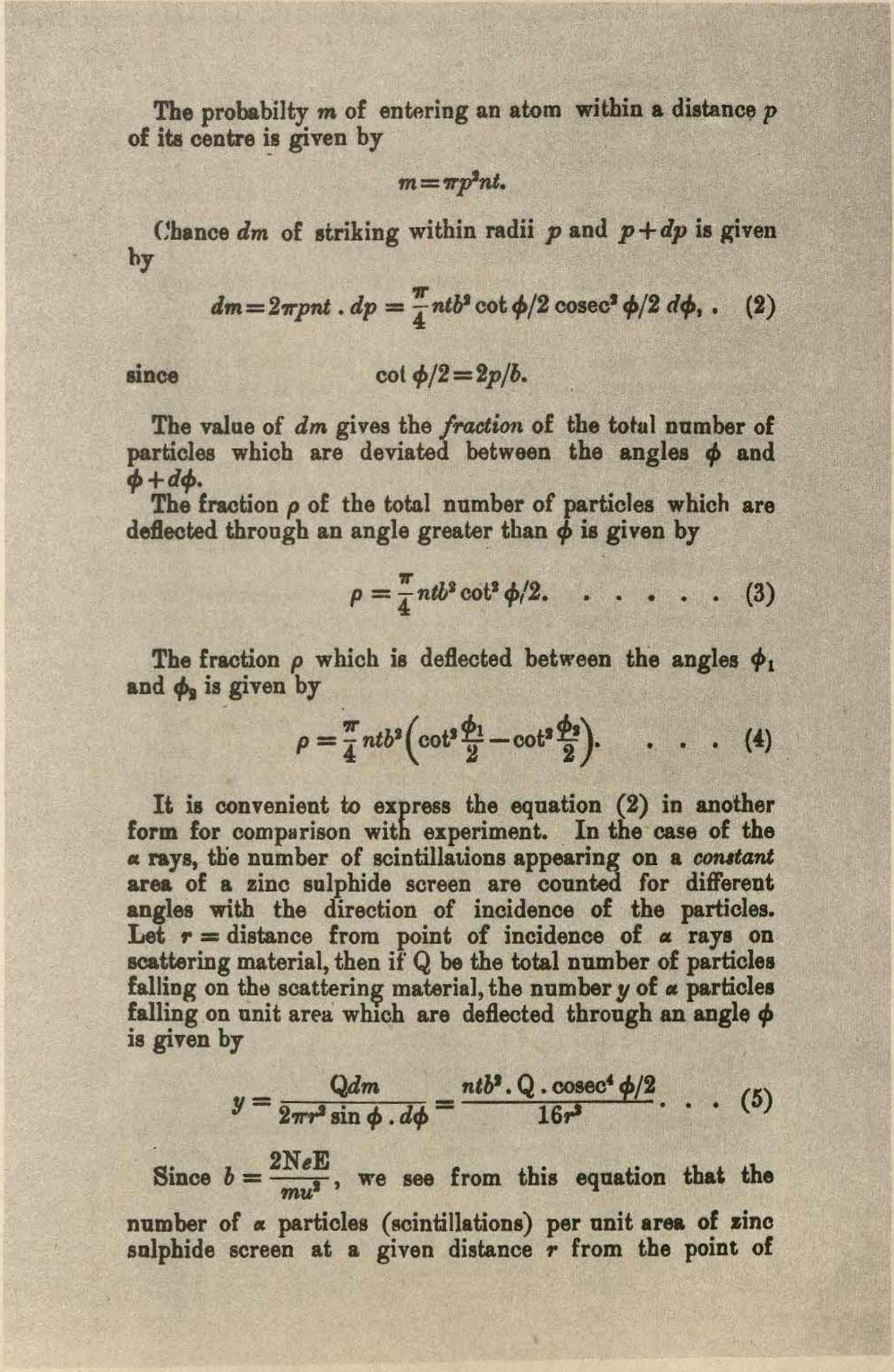
p/b....	10	5	2	1	·5	·25	·125
ϕ......	5°·7	11°·4	28°	53°	90°	127°	152°

§ 3. *Probability of single deflexion through any angle.*

Suppose a pencil of electrified particles to fall normally on a thin screen of matter of thickness t. With the exception of the few particles which are scattered through a large angle, the particles are supposed to pass nearly normally through the plate with only a small change of velocity. Let n=number of atoms in unit volume of material. Then the number of collisions of the particle with the atom of radius R is $\pi R^2 nt$ in the thickness t.

---

* A simple consideration shows that the deflexion is unaltered if the forces are attractive instead of repulsive.



The probabilty m of entering an atom within a distance p of its centre is given by

$$m=\pi p^2 nt.$$

Chance dm of striking within radii p and $p+dp$ is given by

$$dm=2\pi pnt\,.\,dp=\frac{\pi}{4}ntb^2\cot\phi/2\operatorname{cosec}^2\phi/2\,d\phi,\,. \quad (2)$$

since $\qquad \cot\phi/2=2p/b.$

The value of dm gives the *fraction* of the total number of particles which are deviated between the angles ϕ and $\phi+d\phi$.

The fraction ρ of the total number of particles which are deflected through an angle greater than ϕ is given by

$$\rho=\frac{\pi}{4}ntb^2\cot^2\phi/2. \quad . \quad . \quad . \quad . \quad . \quad (3)$$

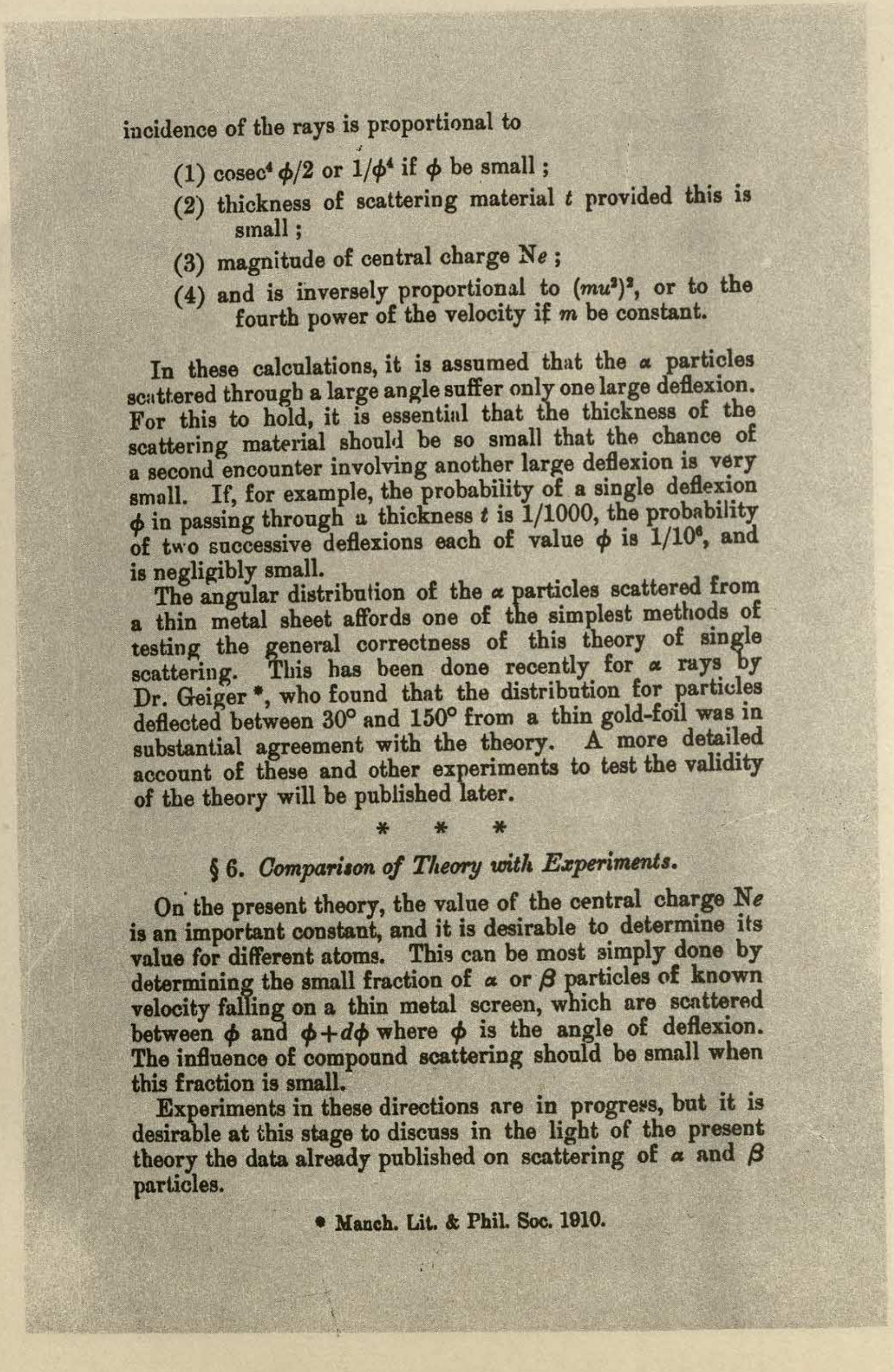
The fraction ρ which is deflected between the angles ϕ_1 and ϕ_2 is given by

$$\rho=\frac{\pi}{4}ntb^2\left(\cot^2\frac{\phi_1}{2}-\cot^2\frac{\phi_2}{2}\right). \quad . \quad . \quad . \quad (4)$$

It is convenient to express the equation (2) in another form for comparison with experiment. In the case of the α rays, the number of scintillations appearing on a *constant* area of a zinc sulphide screen are counted for different angles with the direction of incidence of the particles. Let r = distance from point of incidence of α rays on scattering material, then if Q be the total number of particles falling on the scattering material, the number y of α particles falling on unit area which are deflected through an angle ϕ is given by

$$y=\frac{Q\,dm}{2\pi r^2\sin\phi\,.\,d\phi}=\frac{ntb^2\,.\,Q\,.\operatorname{cosec}^4\phi/2}{16r^2}. \quad . \quad . \quad (5)$$

Since $b=\dfrac{2NeE}{mu^2}$, we see from this equation that the number of α particles (scintillations) per unit area of zinc sulphide screen at a given distance r from the point of



incidence of the rays is proportional to

(1) cosec^4 $\phi/2$ or $1/\phi^4$ if ϕ be small;
(2) thickness of scattering material t provided this is small;
(3) magnitude of central charge Ne;
(4) and is inversely proportional to $(mu^2)^2$, or to the fourth power of the velocity if m be constant.

In these calculations, it is assumed that the α particles scattered through a large angle suffer only one large deflexion. For this to hold, it is essential that the thickness of the scattering material should be so small that the chance of a second encounter involving another large deflexion is very small. If, for example, the probability of a single deflexion ϕ in passing through a thickness t is 1/1000, the probability of two successive deflexions each of value ϕ is $1/10^6$, and is negligibly small.

The angular distribution of the α particles scattered from a thin metal sheet affords one of the simplest methods of testing the general correctness of this theory of single scattering. This has been done recently for α rays by Dr. Geiger *, who found that the distribution for particles deflected between 30° and 150° from a thin gold-foil was in substantial agreement with the theory. A more detailed account of these and other experiments to test the validity of the theory will be published later.

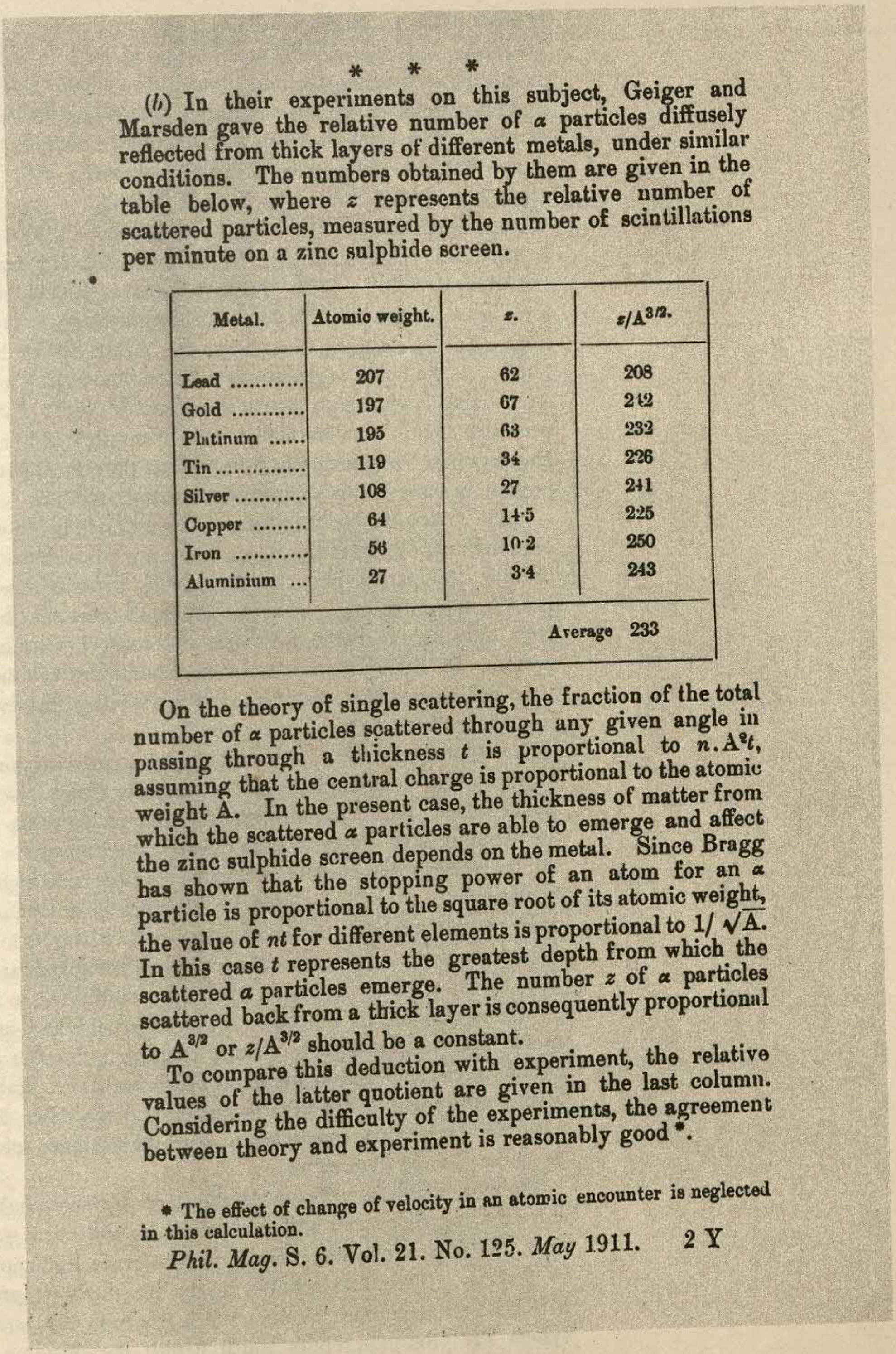
* * *

§ 6. *Comparison of Theory with Experiments.*

On the present theory, the value of the central charge Ne is an important constant, and it is desirable to determine its value for different atoms. This can be most simply done by determining the small fraction of α or β particles of known velocity falling on a thin metal screen, which are scattered between ϕ and $\phi+d\phi$ where ϕ is the angle of deflexion. The influence of compound scattering should be small when this fraction is small.

Experiments in these directions are in progress, but it is desirable at this stage to discuss in the light of the present theory the data already published on scattering of α and β particles.

* Manch. Lit. & Phil. Soc. 1910.



* * *

(*b*) In their experiments on this subject, Geiger and Marsden gave the relative number of α particles diffusely reflected from thick layers of different metals, under similar conditions. The numbers obtained by them are given in the table below, where z represents the relative number of scattered particles, measured by the number of scintillations per minute on a zinc sulphide screen.

Metal.	Atomic weight.	z.	$z/A^{3/2}$.
Lead	207	62	208
Gold	197	67	242
Platinum	195	63	232
Tin	119	34	226
Silver	108	27	241
Copper	64	14·5	225
Iron	56	10·2	250
Aluminium ...	27	3·4	243
		Average	233

On the theory of single scattering, the fraction of the total number of α particles scattered through any given angle in passing through a thickness t is proportional to $n.A^2t$, assuming that the central charge is proportional to the atomic weight A. In the present case, the thickness of matter from which the scattered α particles are able to emerge and affect the zinc sulphide screen depends on the metal. Since Bragg has shown that the stopping power of an atom for an α particle is proportional to the square root of its atomic weight, the value of nt for different elements is proportional to $1/\sqrt{A}$. In this case t represents the greatest depth from which the scattered α particles emerge. The number z of α particles scattered back from a thick layer is consequently proportional to $A^{3/2}$ or $z/A^{3/2}$ should be a constant.

To compare this deduction with experiment, the relative values of the latter quotient are given in the last column. Considering the difficulty of the experiments, the agreement between theory and experiment is reasonably good *.

---

* The effect of change of velocity in an atomic encounter is neglected in this calculation.

Historical Note 2. Bubble chambers

We can get an impression of experimental research in a forefront area of physics by considering the operation of the bubble-chamber system and associated digital computers used in the discovery of many of the unstable particles. This material draws on unpublished portions of a draft of a paper, "Strongly Interacting Particles," written in 1963 by Geoffrey F. Chew, Murray Gell-Mann, and A. H. Rosenfeld.

The bubble chamber can detect to within about 1/1000 in. the track of an electrically charged particle. It was invented in 1952 by Donald Glaser and works as follows: When the particle passes through matter, it ionizes some nearby atoms and transfers some kinetic energy to the recoiling electrons. As the electrons slow down, this kinetic energy heats the liquid locally. If the liquid is already superheated and is looking for a place to start boiling, it will start to boil at these local hot spots. These bubbles are permitted to grow for a few milliseconds; then a light is flashed and they are photographed simultaneously from several different angles so that their position in space can be reconstructed stereographically.

In the earliest bubble chambers, the superheated liquid was a few cubic centimeters of a hydrocarbon which boiled at a convenient temperature and pressure. But particle physicists like to study interactions with the simplest possible target—a single nucleon. Therefore they run their beams into large quantities of liquid hydrogen and study interactions with single protons. By 1955, physicists at several laboratories had developed quart-sized hydrogen-bubble chambers, and it was clear that even larger ones could be made to work.

Next came the problem of interpreting and measuring the tracks. The hydrogen with its insulating vacuum system is always placed in a strong magnetic field which bends the path of the charged particles. By measuring the curvature of the tracks, one can calculate the momentum of the particle. But the highest magnetic fields attainable can only bend high-energy tracks by the order of 10°. To get adequate resolution of momentum (and hence of energy) it is necessary to measure these small curvatures to a few percent. This means we have to measure the position of the photographic image of a track to within a few microns, on a film a few centimeters in size. We need an accuracy of one part in 10,000. The measurements have to be performed rapidly and reliably, because a single chamber a few feet in diameter is capable of yielding 100,000 interesting events each year. Each event might involve measuring five tracks in two or three stereo views; this adds up to 1,000,000 track measurements/year. The old-fashioned microscope must be automated and speeded up.

The most strenuous effort to develop large hydrogen-bubble chambers and data-analysis systems adequate to keep up with them was undertaken by L. W. Alvarez and his group at the Lawrence Radiation Laboratory in Berkeley. Their first solution to the measurement problem was a $100,000 servo-controlled monster which has been very successful. The system has projection optics which permit the operator to view the event as reprojected from the photographs. Once the cross hairs on its optic axis are placed on the track, the servos permit the operator to "drive" automatically down the track (as easily as a locomotive engineer drives a train), meanwhile recording automatically the stage coordinates on IBM cards or punched paper tape. Operated about 120 hours/week, it measures about 30,000 events annually.

The computer programs to analyze the output were written by physicists from several universities. There are three phases to the analysis: spatial reconstruction of individual tracks; kinematic analysis of vertices and whole events; and, finally, statistical analysis of whole experiments. Track reconstruction is a straightforward application of stereo techniques. The reconstruction program calculates the direction and momentum of each track and the uncertainties and correlations in these quantities. The kinematic-event analysis program, called "Kick," is tailor-made for particle physics.

By trying many hypotheses about a particle reaction, Kick can usually sort them out, or at least greatly reduce the number of ambiguities. The track-reconstruction and Kick programs are roughly 10,000 computer instructions in length and are written for the IBM series of 704, 709, 7090, etc., computers. They process a typical event in about four seconds on a 7090. Bubble-chamber data processing now keeps several such computers in the United States and Europe busy full time. Such a computer rents commercially for about $1 million/year.

The size of hydrogen-bubble chambers brought into operation all over the world grew quickly between 1955 and 1959, when a 72-in.-long chamber went into operation at the Bevatron in Berkeley. By early 1960, the measurement rate there was up to one event every few minutes, and the programs were finally working and able to keep up. This capacity was enough to handle the current experiments and to go back and measure some film taken a year earlier.

Bubble-chamber systems are now operating at all major accelerators and have between 1961 and 1963 ploughed up about a dozen new particles. More than half of the particles had all their quantum numbers sorted out by 1963; others were then at various stages of identification. All but one of them are unstable, but it is significant that as late as 1961, a new stable meson, the η singlet was turned up by physicists at Johns Hopkins/Northwestern, working with film produced at the Bevatron. In late 1963, the Berkeley 72-in. chamber was surpassed by an 80-in. chamber brought into operation at Brookhaven National Laboratory.

The whole development process, from the invention of the bubble chamber (1952) to production-line discovery of particles (1961), took nearly 10 years. This illustrates the complexity of particle physics. In a similar way, the discovery of the antiproton in 1955 at the Bevatron hinged on the 1948 decision to build the Bevatron, which was the first accelerator designed to give protons enough energy to create antiprotons artificially. We see that ten years is not a long lead time for a large technological project.

Index

A

- Aberration, of light, 350–351
  - of starlight, 313–316
- Ablation, meteoroid, 189
- Accelerated clocks, 363–364
- Acceleration, 28, 83–86
  - absolute and relative, 63–64
  - angular, 181–182
  - centripetal, 31, 59, 76
  - in circular motion, 31, 81
  - Coriolis, 84–86
  - cyclotron, 109–110
  - gravitational, 292
    - (*See also* g)
  - longitudinal, of electron, 101
    - of proton, 100–101
  - relativistic, 401, 408–413
  - transverse, 101–102
- Accelerators, particle, 448
  - Van de Graaff, 337
- Accelerometer, 76
- Addition of velocities, relativistic, 350
- Age of universe, 328
- Alvarez, L. W., 468
- Amplification, parametric, 229–231
- Amplitude, 212
- Angular frequency (ω), 30, 195
- Angular momentum (**J**), 175
  - conservation of, 175–180
  - of elementary particles, 185
  - of moon, 186
  - in solar system, 184–185
  - spin, 177, 185
- Angular velocity (ω), 30, 89–90
- Anharmonic oscillator, 197, 227–229
- Anticyclic order, 41
- Antineutron ($\bar{n}$), 441
- Antiproton ($\bar{p}$), 441
- Antiproton threshold, 417–418
- Area of parallelogram, 38
- Aslakson, C. I., 320
- Astronomical unit of length (A.U.), 289
- Axes, principal, 249

B

- Baryons, 446, 456
- Bergstrand, E., 312
- Bertozzi, W., 337
- Beta emission, 444
- Bevatron, 417
- Binary star, 292
- Binding energy, of deuteron, 392
  - of nuclei, 392
- Binomial theorem, 158
- Bohr, N., 440
- Bohr radius, 272
- Bohr-Rutherford atom, 440
- Boltzmann constant (k), 300
- Boundary conditions, 100
- Bradley, James, 313–315
- Bridgman, P. W., 62
- Brookhaven National Laboratory, 443, 448
- Bubble chamber, 452–453, 468–469

C

- *c*, constant, 311–312
- Capacitance, 205
- Cartesian coordinate system, 40–42, 52–53, 242
- Cavity resonator, 318
- Cells of living matter, 3–6
- Center of mass, 167–168
- Center-of-mass reference frame, 169, 172, 413, 416
- Central force, 145, 176
- Centrifugal force, 76
- Centripetal acceleration, 31, 59, 76
- Cepheid stars, 340

Čerenkov effect, 311
Ceres, discovery of, 162
CERN (Conseil européen pour la recherche nucléaire), 442, 448
CGS units, 94
Chadwick, J., 441
Charge, elementary, 95
Charge invariance, 402
Charged particles, 451
  in constant electric field, 99–100, 408–410
  in magnetic field, 106, 411–413
  in uniform alternating electric field, 102–104
Chemical reactions, 73
Circuit, *LC*, 204
  LCR, 223
Circular hoop, rotation of, 238
Circular motion, 30, 76
Clocks, accelerated, 363–364
  time dilation, 356–361
Cogwheel apparatus, Fizeau, 318
Collision, of atom and molecule, 73
  kinematics, 82–83
  of particles, 73–75, 186–187, 446–447
    energy from, 415–417
    heavy and light, 74–75
    with internal excitations, 170
    sticking together, 168–170
  relativistic dynamics, 389–390
Comet, 399
  radiation pressure in tail, 398–399
Complex conjugate, 119
Complex numbers, 118–126, 231–233
Complex variables, 225
Compton wavelength, 15, 258, 271
Condensation of galaxies, 301
Conservation, of energy, 70, 133–163
    law of, 138, 140
  of momentum, 69–73, 165–189, 382–385
    angular, 175–180
    linear, 73, 165–167
Conservation laws, 69, 133–134, 382
Conservative forces, 145–147
Contraction, length, 354
Coriolis acceleration, 84–86
Cosines, direction, 35
  law of, 34
Cosmic-ray particles, 402
Coulomb force, 261
Coulomb's law, 94
Crystal lattices, 49–50
Curvature of space, 9–11
Cyclic order, 41
Cycloid, 115–116
Cyclotron, 109–110, 113, 127–128
  relativistic, 411
Cyclotron frequency, 106, 413
Cyclotron radius, 106
Cylindrical polar coordinates, 48

D

Damped harmonic oscillator, 207–209, 223
Damping force, 206
De Moivre equation, 121
Deoxyribonucleic acid (DNA), 4, 6
Derivatives, vector, 28
Determinants, 42, 47
Deuteron, binding energy of, 392
Dicke, R., 430
Differential equations, 99, 126
Differentiation, of cosine, 30
  partial, 151
  of products of vectors, 89
  of sine, 30
  of vectors, 28
Dilation, time, 356–361
Dimensionless numbers, 271
Dirac, P. A. M., 6, 41
Direction cosines, 35
Direction of winds, 88
Disintegration energies, 394
Dispersive medium, 312
Distributive law, 37
DNA (deoxyribonucleic acid), 4, 6
Doppler effect, 322, 324–325, 327–328, 339, 366
  longitudinal, 361–362
  relativistic, 362, 366
  transverse, 362
Doppler shift (*see* Doppler effect; Red shift)
Driven harmonic oscillator, 211–214, 223, 231–233

Durbin, R. P., 358
Dushman, L., 394
Dwarf galaxy, 293
Dynamics, relativistic, 381–425
  of rigid bodies, 235–259

E

Earth, center of, 293
  wobble of, 251
Eccentricity (e), 283–285
Eclipses of Io, satellite of Jupiter, 312, 339
Ecliptic, 15
Effective centrifugal potential energy, 187
Einstein, Albert, 12, 375, 404
  general theory of relativity, 374
  principle of equivalence, 435
  special theory of relativity, 375–379
Electric charge, 94
Electric field intensity (**E**), 95
Electrodynamics, laws of, 5
Electromagnetic waves, 36
Electron(s), in bound orbit, 160
  classical radius of, 270
  in oscilloscope, 162
Electron-positron pair, 413
Electron volts (ev), 154–155
Electrostatic force, 96, 261
Electrostatic potential, 152–154, 262
Electrostatic self-energy, 267
Elementary charge (e), 95
Elementary particles, 444
Elevator, freely falling, 77
Ellipse, 283
Elliptic integral, 227
Energy, conservation of, 70, 133–163
  equipartition of, 300
  of harmonic oscillator, 201
  kinetic (*see* Kinetic energy)
  potential (*see* Potential energy)
  relativistic, 386–387
  solar, 403
  stellar, 294, 394
  threshold, 413–415
Energy function, 138
Energy levels, 277
Eötvös, R., 428
Equilibrium, static, 274
Equipartition of energy, 300
Equipotential surfaces, 274
Equivalence, 427–437
  Einstein principle of, 435
  of mass and energy, 390–394, 404–405
Erg, 136
Escape of molecules, 160
Escape velocity, 156–157, 161
Ether, 330, 376
Euclidean geometry, 7–9, 12–13
Euler equations, 250

F

Feynman, R. P., 6, 305
Fictitious forces, 75
Fields, crossed, 114
  electric, 95, 97
  magnetic, 97, 105
Fine-structure constant, 311
Fitzgerald contraction, 355
Fizeau's cogwheel apparatus, 318
Force, central, 145
  centrifugal, 76
  conservative, 145–147
  Coriolis, 84–86
  damping, 206
  electrostatic, 96
  fictitious, 75
  gravitational (*see* Gravitational force)
  inverse-square-law, 94, 261–309
  on ladder, 186
  linear restoring, 143–144, 155–156
  Lorentz, 98, 146
  magnetic, 39, 97
  relativistic, 400
Force constant, 143
Form invariant, 41
Foucault pendulum, 77–79
Franklin, Benjamin, 439
Free fall, 86–87, 142–143
Freely falling elevator, 77

Frequency, angular, 30, 195
  fundamental, 199
Friction, 205–207
Fundamental frequency, 199
Fundamental length, 270

G

g, centrifugal correction of, 88
  determination of, 294
  values of, 59
Galaxy (galaxies), 60, 64, 183
  condensation, 301–302
  dwarf, 293
  gravitational energy, 268
  motion in, 292–293
  origin, 303–305
  rotation, 340
  shape, 182–183
  velocity, 341, 365
Galilean invariance, 64–65, 70–73
Galilean transformation, 66–69, 328–329, 333
  failure of, 346–348
Galileo, 61
Gamma rays, 342–343
  recoil, 415
Gauss, C. F., 8–9, 97
Gaussian units, 93–98
Generator, definition of, 180*n*.
Genetics, DNA molecule in, 5–6
Geodimeter, 322
Geometry, 7–13
Gibbs, J. W., 26, 53
Gimbal mounting, 257
Glaser, D., 453
Globular clusters, 294
Granularity of matter, 439
Gravitation, law of, 61, 63, 79–80
Gravitational energy, of Galaxy, 268
  of sphere, 268–270
Gravitational field, 224
Gravitational force, 94, 261, 291–294
Gravitational length, 272
Gravitational mass, 427–430
  of photons, 430–431
Gravitational potential, 157–158, 262
Gravitational red shift, 433–434
Gravitational self-energy, 267
Greenstein, J. L., 437
Gyrocompass, 257
Gyrofrequency, 106, 108
Gyroradius, 106, 108
  relativistic, 413
Gyroscope, 235, 255–257
Gyroscopic stabilization, 259

H

Hadamard, J., 314
Harmonic oscillator, 151, 191–233, 282
  damped, 207–209, 223
  driven, 211–214, 223, 231–233
  energy, 201
  power absorption, 217–218
  power dissipation, 209–211
Havens, W. W., Jr., 358
HCl (hydrogen chloride), 278
Heaviside, O., 26
Helix, 106
HF (hydrogen fluoride), 278
Hooke's law, 143, 200
Hoop, circular, 238
Hydrogen, burning of, 394
Hydrogen atom, Bohr radius of, 272
Hydrogen chloride, 278
Hydrogen fluoride, 278

I

Impulse, 135
Impulse approximation, 435
Inclined plane, 248
Inductance, 204
Inertia, 238
  moment of (*see* Moment of inertia)
Inertial coefficients, 240–241
Inertial mass, 427–430
Inertial reference frames, 58–63, 328–331
Initial condition, 100
Integral, line, 141

Interaction, strong, 457
  weak, 457
Internal torques, 177
International Latitude Service, 251
Invariance, 13
  charge, 402
  Galilean, 64–65, 70–73
  rotational, 13, 180
  of velocity of light (c), 336
Inverse-square-law force, 94, 261–309
Inversion of coordinate axes, 46
Io, eclipses of, 312, 339
Ions, high-speed, 128–130
Ives, H. E., 362

J

Joule, 136
Jupiter, eclipses of Io by, 312, 339

K

Kepler problem, 274
Kepler's laws, 178, 288–290
Kerr cell, 320
Kinetic energy, 135, 141–142, 159, 160
  of harmonic oscillator, 201–202
  of rotation, 246–248
Klein, F., 235
Kronecker delta symbol (δ_{ij}), 52, 246

L

Ladder, forces on, 186
Larmor, J., 349
Lattices, crystal, 49–50
Law(s), of classical mechanics, 5
  of cosines, 34
  of electrodynamics, 5
  of physics, 6–7
  of quantum mechanics, 5
  of sines, 38
  (*See also* Kepler's laws; Newton's laws)
Lawrence, E. O., 110, 127, 128
LC circuit, 204
LCR circuit, 223
Length, contraction of, 354
  fundamental, 270
  measurement of, 353
Leptons, 446, 456
Levi-Civita tensor, 52
Lifetime of particles, 454–456
Light, curved path of, 11–12
  speed of, 311–343
Light cone, 369
Line integral, 141
Linear momentum, conservation of, 73, 165–167
Linear response, 191
Linear restoring force, 143–144, 155–156
Livingston, M. S., 110, 127, 128
Loar, H. H., 358
Local Group, 60
Longitudinal acceleration, 100–101
Longitudinal Doppler effect, 361
Lorentz force, 98, 146
Lorentz-invariant, 364, 386
Lorentz transformation, 345–379, 388

M

McDonald, D. F., 312
Mach's principle, 64
McMillan, E. M., 422, 423
McVitties, G. C., 328
Magnetic field, 97, 105
Magnetic field intensity, 97
Magnetic focusing, 109
Magnetic force, 97
Magnetic monopoles, 113
Mass, gravitational, 427–431
  inertial, 427–430
  of particle, 446, 448–450
  of photon, 430
  reduced, 275–277
  relativistic, 385–386
  rest, 382, 388
  on spring, 200
  of sun, 15

Mass defect, 393
Mass-energy conversions, 391–394
Mass-energy equivalence, 390–394, 404–405
Mass spectrometer, 161
Matrix, 241
Matter, inanimate, 4
  living, 3–6
Maxwell, J. C., 330
Mean life of particles, 446, 448–450, 454
Mercury, orbit of, 12
Mesons, 456
  μ (mu), 365
  μ^+, 358, 441
  π (pi), 441, 450
  π^+, 358, 441
Meteoroids, encounter with atmosphere, 188–189
  orbits, 293
Michelson, A. A., 318, 332
Michelson-Morley experiments, 332–336, 345
Millikan oil-drop experiment, 223
Minkowski, H., 366
MKS units, 98, 136
Modulated light detector, 320
Modulation, 229
  square-wave, 231
Modulus, 119
Molecules, escape of, 160
  vibration of, 277–279
Moment of inertia, 238–240, 245, 258
  of cylinder, 245
  of sphere, 243–245
  of spherical shell, 243
Momentum, 135
  angular (*see* Angular momentum)
  conservation of, 69–73, 165–189, 382–385
    linear, 73, 165
  relativistic, 382–385
Momentum-energy transformation, 388
Momentum selector, 109
Monopoles, magnetic, 113
Moon, angular momentum, 186
  escape velocity from, 161
Morley, E. W., 332
Mössbauer effect, 342–343
Mu mesons (*see* Mesons)
Mulligan, J. F., 312

N

Neptune, angular momentum of, 184
  discovery of, 10, 162
Neutrino, 345, 395, 441
Neutron, charge distribution, 451
Neutron decay, 403
Neutron stars, 273
Newcomb, Simon, 253
Newton pendulum, 428, 437
Newtonian force, 70
Newton's laws, of gravitation, 61, 63, 79–80
    rotating bucket example, 63–64, 90–91
  of motion, 55–58
Newtons (units), 136
Newtons/coulomb, 98
Nonlinear pendulum, 197
Nonrelativistic dynamics, 93–131
Novae, 341
Nuclear reactions, 302
Nuclei, binding energy, 392
Nucleon-nucleon interaction, 161
Numbers, dimensionless, 271
  fundamental, 270

O

Ohm's law, 206
One-body problem, 279–283
Orbits, of meteoroids, 293
  of particles, 274–275, 283
    circular, 286–287
    eccentricity, 283, 285
  of planets, 12, 178–179, 288–290
    numerical calculation, 305–309
  of satellites (*see* Satellites)
  of stars, 292
Orthogonal axes, 35

Oscillator, 197
  anharmonic, 197, 227–229
  harmonic (*see* Harmonic oscillator)
Oscilloscope, 162

P

Packing-fraction curve, 393
Parallax, 10–11
  stellar, 340
Parallel-axis theorem, 244–245
Parallelepiped, volume of, 38
Parallelogram, area of, 38
Parallelogram law of addition, 27
Parametric amplification, 229
Parsec, 289
Partial derivatives, 151
Particle accelerators, 448
Particles, 439–469
  charge of, 451
  charged (*see* Charged particles)
  collision of (*see* Collision)
  deflection of, 170–173
  lifetime of, 454–456
  masses and mean lives of, 446, 448–450, 454
  orbits of, 274–275, 283, 286–287
  in rigid body, 235–237
  speed of, 28
  spin angular momentum of, 185
  with zero rest mass, 397–401
Pendulum, 192–196
  exact solution, 225–227
  Foucault, 77–79
  Newton, 428, 437
  nonlinear effects, 197–199
Period, 32, 80
  satellite, 80
Phase amplitude, 194
Phase angle, driven oscillator, 213
Phase plane, 226
Photon deflection by sun, 434–435
Physical laws, 6–7
Pic du Midi, 442
Pions, 358, 441, 450
Planck's constant, 271, 311, 404
Planck's length, 294
Plane, equation of, 35
Planets, angular momentum of, 184
  discovery of, 10, 162–163
  orbits of, 12, 178–179
    Kepler's laws, 288–290
    numerical calculation, 305–309
Pluto, discovery of, 10
Polar coordinates, 48
Polar plot, 214
Positron, 441
Positronium, 277
Potential, 153
  effective, 281
  electrostatic, 152
  gravitational, 157–158, 262
  Yukawa, 161
Potential difference, 152
Potential energy, 136–137, 147–153, 159–160
  effective centrifugal, 187
  in electric field, 151–153
  electrostatic, 159
  of harmonic oscillator, 201–202
  inverse-first-power law of, 262
Pound, R. V., 431
Power, 144–145
Power absorption, harmonic oscillator, 217–218
Power dissipation, harmonic oscillator, 209–211
Precession of sphere, 251
Pressure, radiation, 398
Principal axes, 249
Proper time, 356
Proton, charge distribution, 451
  in crossed electric and magnetic fields, 114–115
Proton scattering by heavy nucleus, 179–180
Proton synchrotron, 443
Pseudoforce, 75
Psychology of invention, 314
Pythagorean theorem, 7, 192

Q

Quality factor (Q), 211, 224
Quantum mechanics, laws of, 5

R

Radar beacon, 319
Radians, 30
Radiation pressure, 398
Radiative energy losses, 302
Radius, of electron, 270
  of universe, 11, 272, 328
Rebka, S., Jr., 431
Recessional red shift, 327
Recessional velocity of galaxy, 365
Reciprocal lattice, 49
Recoilless emission, 342
Red shift, gravitational, 433–434
  recessional, 327
  (*See also* Doppler effect)
Reduced mass, 275
Reference frames, center-of-mass, 169–170, 172, 413, 416
  inertial, 58–63, 328–331
  noninertial, 83
  transformations of, 116–118
Refractive medium, 312
Relativistic acceleration, 401, 408–413
Relativistic cyclotron, 411
Relativistic dynamics, 381–425
Relativistic energy, 386–387
Relativistic mass, 385–386
Relativistic momentum, 382–385
Relativistic rocket equation, 418
Relativity, general theory of, 374
  special theory of, 345, 375–379
Relaxation time, 205
Resistor, 206
Resonance, 216, 232
  spin, 254–255
Rest mass, 382, 388
  zero, 397–401
Reynolds' number, 271
Right-hand-thread rule, 36–37
Rigid bodies, 235–259
  equations of motion, 238
  particles in, 235–237
  rotation of, 33, 238–243
Rocket, relativistic equation, 418
Roemer, O., 312
Rolling sphere, 248
Rotating bucket, 63–64, 90–91
Rotating coordinate systems, 83
Rotating mirrors, 318
Rotation, of galaxies, 340
  kinetic energy of, 246–247
  of rigid body, 33, 238–243
Rotational invariance, 180–181
Rutherford, E., 440
Rutherford scattering, 179, 459–467

S

Sadeh, D., 352
Satellites, angular momentum of, 185–186
  deceleration of, 173
  in orbit, 80, 160
    period of, 80
  propulsion equation, 174
  speed of, 173
Scalar, 26
Scalar field, 44
Scalar product, 33–36
Scattering, proton, 179–180
  Rutherford, 179, 459–467
Schiff, L. I., 435
Schwarzschild, K., 10
Sea, formation of, 294–296
Self-energy, 267–268
Series expansion, 158
Shoran, 319
Simultaneity, 356, 365, 370, 376
Sines, law of, 38
Sirius B, 292
Solar energy, 294, 403
Solar system, angular momentum, 184
Sommerfeld, A., 235
Space-time, 366–375
Space vehicle, 173–175
Speed, of light (c), 311–343, 345
  ultimate, 337
Sphere, self-energy, 268
Spherical polar coordinates, 48
Spin angular momentum, 177, 185
Spin precession, 253–254
Spin resonance, 254–255
Spring, 143
Square-wave modulation, 231

Starlight, aberration of, 313–316
Stars, binary, 292
  Cepheid, 340
  gravitational problems, 292
  neutron, 273
  origin of, 303
  variable, 340
  white dwarf, 273, 292
Statcoulomb, 94
Statvolt, 95, 152–153
Statvolt/cm, 95
Stellar energy, 294, 394–395
Stellar parallax, 340
Stilwell, G. R., 362
Stokes' law, 225
Strong interaction, 457
Summation convention, 245, 259
Sun, diameter and mass, 15–17
  photon deflection, 434
  temperature, 300–301
Superposition, 191
Superposition principle, 219–220
Symmetrical top, 251–252
Synchrotron, 422–425
  proton, 443

T

Temperature of sun, 300–301
Tensor, 25–26
Thermal expansion, 229
Third harmonic, 199
Thompson, J. J., 439
Threshold energy, 413–415
  of antiproton, 417
Tides, origin of, 296
Time derivative, 28
Time dilation, 356–361
Torque, 47, 175, 251
  internal, 177
Torricelli, E., 330
Torsion suspension, 224
Transformation, Galilean (*see* Galilean transformation)
  Lorentz, 346
  of reference frame, 116–118
  relativistic, 388
Transponder, 319
Transverse Doppler effect, 362
Transverse momentum components, 170
Trigonometrical parallax, 10–11
Triple product of vectors, 39
Two-body problem, 275–277

U

Ukiah, Calif., 251
Ultimate speed, 337
Ultracentrifuge, 56–58
Universe, age of, 328
  radius of, 11, 272, 328
  size of, 3
Uranus, 163, 290

V

Van de Graaff accelerator, 337
Variable stars, 340
Vector(s), 25–53
  addition, 27
  derivatives, 28–33
  identities, 44
  notation, 25–26
  products, 33–39, 47, 89
    triple, 39
Vector field, 44
Veksler, V., 422
Velocity, 28, 83–86, 89–90
  absolute and relative, 64
  angular, 30, 89–90
  escape, 156–157, 160
  in viscous medium, 223
Velocity addition, 350–353
  relativistic, 350
Velocity distribution function, 160
Velocity selector, 114
Vibration of diatomic molecule, 277–279
Vibrational frequencies, 278
Virial theorem, 296–299
Viscosity, 225
Viscous medium, 223

Volt, 152–153
Voltage drop, 152
Volume of parallelepiped, 38

W

Watt, 145
Weak interaction, 457
Webers/m^2, 99
Weisskopf, V. F., 355
Wheeler, J. A., 366
White dwarf star, 273, 292
Winds, direction of, 88
Witten, L., 436
Wobble of earth, 251
Work, 135, 141
  rate of doing, 36
  relativistic energy and, 395–39
World line, 371

Y

Yukawa potential, 161

Z

Zeeman, P., 40
Zeeman effect, 40, 439
Zenith, 15
Zero rest mass, 397–401

Table of Values

Value and Units	*Item*	*Symbol or Abbreviation*	*Derivation of Value*
• *Gases*			
22.4×10^3 cm^3/mole	Molar volume at S.T.P.	V_0	
2.69×10^{19} cm^{-3}	Loschmidt's number	n_0	N_0/V_0
6.0225×10^{23} mole^{-1}	Avogadro's number	N_0	
8.31×10^7 erg mole^{-1} deg^{-1}	Gas constant	R_0	
1.381×10^{-16} erg/deg K	Boltzmann's constant	k, k_B	R_0/N_0
1.01×10^6 dyne/cm^2	Atmospheric pressure		
$\sim 10^{-5}$ cm	Mean free path of nitrogen at S.T.P.		
3.32×10^4 cm/sec	Speed of sound in air at S.T.P.		
• *Atomic*			
6.626×10^{-27} erg-sec	Planck's constant	h	
1.054×10^{-27} erg-sec	Planck's constant/2π	$\hbar$	$h/2\pi$
13.6 electron volts	Energy associated with 1 Rydberg		
1.98×10^{-16} erg	Energy associated with unit wave number		hc
1.60×10^{-12} erg	Energy associated with 1 electron volt	ev	
1.24×10^{-4} cm	Wavelength associated with 1 electron volt	λ_0	hc^2/e
8066 cm^{-1}	Wave number associated with 1 electron volt		$1/\lambda_0$
2.42×10^{14} sec^{-1}	Frequency associated with 1 electron volt	ν_0	c/λ_0
0.529×10^{-8} cm	Bohr radius of the ground state [illegible] hydrogen	a_0	$\hbar^2/me^2$
$\sim 10^{-8}$ cm	Radius of an atom		
0.927×10^{-20} erg/gauss	Bohr magneton	μ_B	$e\hbar/2mc$
0.505×10^{-23} erg/gauss	Nuclear magneton	μ_N	$e\hbar/2M_pc$
137.04	Reciprocal of fine-structure constant	α^{-1}	$\hbar c/e^2$
• *Particles*			
1.6725×10^{-24} gm	Proton rest mass	M_p	
1.6747×10^{-24} gm	Neutron rest mass	M_n	
1.66042×10^{-24} gm	1 unified atomic mass unit ($\equiv 1/12$ mass of C^{12})	u	
0.911×10^{-27} gm	Electron rest mass	m	
0.9382×10^9 ev	Energy equivalent to proton rest mass	E_p	M_pc^2
0.9395×10^9 ev	Energy equivalent to neutron rest mass	E_n	M_nc^2
0.511×10^6 ev	Energy equivalent to electron rest mass		mc^2
1836	Proton mass/electron mass		M_p/m
2.82×10^{-13} cm	Classical radius of the electron	r_0	e^2/mc^2
4.803×10^{-10} esu	Charge on the proton	e	
3.86×10^{-11} cm	Electron Compton wavelength	$\lambdabar_{Ce}$	$\hbar/mc$